

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:55:55 ; Search time 16.351 Seconds
(without alignments)
4244.936 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 722

Sequence: 1 MGQFMSVFRIMNTATVLA.....NTFNIGSERTWTANLRYSF 722

Scoring table: OLIGO

Searched: 283224 seqs, 96134422 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 1000 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	172	23.8	708	2	Tomb-dependent rec
2	11	1.5	853	2	ferrichrome-iron r
3	9	1.2	629	2	1-deoxyxy/ulose-5-
4	9	1.2	725	2	probable ferric ai
5	9	1.2	725	2	Tomb-dependent rec
6	8	1.1	284	2	hypothetical prote
7	8	1.1	532	1	59k transcription
8	8	1.1	644	2	probable alpha-iso
9	8	1.1	696	2	ferrinoxamine B rec
10	8	1.1	720	2	Fe(III)-pyochelin
11	8	1.1	753	2	probable iron-side
12	8	1.1	730	2	hypothetical prote
13	8	1.1	813	2	ferrityoverdine re
14	8	1.1	815	2	hypothetical prote
15	8	1.1	837	2	hypothetical prote
16	8	1.1	1274	2	hypothetical prote
17	8	1.1	78	2	hypothetical prote
18	7	1.0	84	2	hypothetical prote
19	7	1.0	102	2	thioredoxin TC0826
20	7	1.0	108	2	hypothetical prote
21	7	1.0	108	2	hypothetical prote
22	7	1.0	112	2	probable secreted
23	7	1.0	123	2	hypothetical prote
24	7	1.0	127	2	60S ribosomal prot
25	7	1.0	127	2	ATP synthase prote
26	7	1.0	130	2	ribosomal protein
27	7	1.0	146	2	transcription regu
28	7	1.0	148	2	ribosomal protein
29	7	1.0	149	1	ribosomal protein

30	7	1.0	155	2	S77384	hypothetical prote
31	7	1.0	166	2	AH3359	probable competent
32	7	1.0	165	2	T10313	hypothetical prote
33	7	1.0	173	2	F83062	hypothetical prote
34	7	1.0	174	2	C86231	hypothetical prote
35	7	1.0	177	1	PMECD	H+-transporting tw
36	7	1.0	177	2	AC0954	ATP synthase delta
37	7	1.0	177	2	F86059	H+-transporting tw
38	7	1.0	177	2	AH0500	H+-transporting tw
39	7	1.0	177	2	E91213	hypothetical prote
40	7	1.0	182	2	D82490	hypothetical prote
41	7	1.0	191	2	T04031	hypothetical prote
42	7	1.0	201	2	E95072	nitroreductase fam
43	7	1.0	201	2	B89740	nitroreductase (BC
44	7	1.0	208	2	B86874	hypothetical prote
45	7	1.0	209	2	AC2841	outer membrane pro
46	7	1.0	211	2	S47347	outer membrane pro
47	7	1.0	214	2	I49351	endotoxin 3 precu
48	7	1.0	217	2	D82407	outer membrane pro
49	7	1.0	219	2	H97618	25k outer-membrane
50	7	1.0	220	2	AE2360	hypothetical prote
51	7	1.0	221	2	T48101	hypothetical prote
52	7	1.0	236	2	G75324	hypothetical prote
53	7	1.0	241	2	T51206	hypothetical prote
54	7	1.0	256	2	S39718	spore coat polysac
55	7	1.0	260	2	AD2584	hypothetical prote
56	7	1.0	260	2	D97366	fica protein (AF19
57	7	1.0	268	2	A70379	conserved hypothet
58	7	1.0	269	2	D84312	sulfate transport
59	7	1.0	275	2	H87538	hypothetical prote
60	7	1.0	280	2	T03942	rRNA N-glycosidase
61	7	1.0	287	2	F75057	hypothetical prote
62	7	1.0	286	2	H64113	hemagglutinin ha1
63	7	1.0	295	2	S76136	hypothetical prote
64	7	1.0	306	2	B85436	RNA-binding like p
65	7	1.0	307	2	AE3487	dieneacetone hydro
66	7	1.0	310	2	T31799	hypothetical prote
67	7	1.0	312	2	A64660	site-specific DNA-
68	7	1.0	318	2	G90707	hypothetical prote
69	7	1.0	318	2	C85558	hypothetical prote
70	7	1.0	318	2	AD0575	ferriterobactin-
71	7	1.0	318	2	UV0045	ferriterobactin-
72	7	1.0	319	2	F69483	conserved hypothet
73	7	1.0	321	2	B71854	type II DNA modifi
74	7	1.0	324	2	C95048	3-oxoacyl-(acyl-ca
75	7	1.0	324	2	A97919	3-oxoacyl-(acyl-ca
76	7	1.0	329	2	T49953	prx10 peroxidase-1
77	7	1.0	330	2	A70422	hypothetical prote
78	7	1.0	331	1	G69830	lipoate-protein 11
79	7	1.0	334	2	H64079	ornithine carbamoy
80	7	1.0	339	2	T25158	hypothetical prote
81	7	1.0	339	2	AE2212	hypothetical prote
82	7	1.0	342	2	AH1071	conserved hypothet
83	7	1.0	344	2	T11620	hypothetical prote
84	7	1.0	351	2	S25520	outer membrane por
85	7	1.0	354	2	C71368	probable UDP-N-ace
86	7	1.0	360	2	C82220	hypothetical prote
87	7	1.0	363	2	F43680	D1363 protein - Af
88	7	1.0	372	2	G83707	hypothetical prote
89	7	1.0	373	1	VB8EKA	glycoprotein gX pr
90	7	1.0	382	1	A60112	42k membrane anti
91	7	1.0	382	2	S15578	ipac protein - Shi
92	7	1.0	382	2	T04347	teosinte branched
93	7	1.0	383	2	T34603	xylanase A - Stre
94	7	1.0	394	2	F95973	probable sugar upt
95	7	1.0	395	2	A43700	BNS1 protein - hum
96	7	1.0	409	2	T47298	probable replicati
97	7	1.0	411	1	VCBEG2	glycoprotein G pre
98	7	1.0	411	2	T01388	oxidoreductase hom
99	7	1.0	412	2	D72467	hypothetical prote
100	7	1.0	421	2	T31787	hypothetical prote
101	7	1.0	421	2	E71363	hypothetical prote
102	7	1.0	423	2	A70947	hypothetical prote

103	7	1.0	426	2	E87321	176	7	1.0	1280	2	G96796	hypothetical prote
104	7	1.0	437	2	C64113	177	7	1.0	1283	2	T13799	neurexin IV - frui
105	7	1.0	443	2	S57328	178	7	1.0	1341	2	JG0166	LaMDRI protein - L
106	7	1.0	454	2	E70448	179	7	1.0	1388	2	T38720	chromodomain helic
107	7	1.0	454	2	T12539	180	7	1.0	1514	2	S70099	hypothetical prote
108	7	1.0	456	2	T11665	181	7	1.0	1545	2	T14288	DNA (cytosine-5)-
109	7	1.0	457	2	H84716	182	7	1.0	1557	2	T29132	hypothetical prote
110	7	1.0	461	2	T00918	183	7	1.0	1599	2	T15854	hypothetical prote
111	7	1.0	477	1	JS0589	184	7	1.0	1616	2	T47801	hypothetical prote
112	7	1.0	477	2	S55494	185	7	1.0	1643	1	RRWGNV	genome polyprotein
113	7	1.0	478	2	T05157	186	7	1.0	1761	2	T14289	DNA (cytosine-5)-
114	7	1.0	479	2	C97508	187	7	1.0	1772	2	T36105	probable large gly
115	7	1.0	479	2	AG2726	188	7	1.0	2014	2	S46622	probable membrane
116	7	1.0	485	2	S48650	189	7	1.0	2164	1	GNNY89	genome polyprotein
117	7	1.0	499	2	T01230	190	7	1.0	2551	2	S98047	hypothetical prote
118	7	1.0	505	1	D70703	191	7	1.0	4572	2	S57908	hypothetical 527K
119	7	1.0	505	2	T51741	192	6	0.8	35	2	H86021	hypothetical prote
120	7	1.0	506	2	T14907	193	6	0.8	39	2	AB3109	hypothetical prote
121	7	1.0	506	2	AE0596	194	6	0.8	44	2	D71666	ribosomal protein
122	7	1.0	506	2	S73862	195	6	0.8	44	2	S36977	hypothetical prote
123	7	1.0	514	2	G81935	196	6	0.8	44	2	H97816	50S ribosomal prot
124	7	1.0	514	2	A81170	197	6	0.8	50	2	G91003	hoat-nuclease inhi
125	7	1.0	518	2	F70687	198	6	0.8	53	2	A82862	hypothetical prote
126	7	1.0	520	2	E87621	199	6	0.8	55	2	A37238	autoimmune epitope
127	7	1.0	534	2	T30019	200	6	0.8	57	2	G98090	hypothetical prote
128	7	1.0	550	2	E85574	201	6	0.8	59	2	D84234	hypothetical prote
129	7	1.0	550	2	S85574	202	6	0.8	60	2	S62076	M-like protein pre
130	7	1.0	554	2	S37936	203	6	0.8	60	2	D82277	hypothetical prote
131	7	1.0	558	1	S68203	204	6	0.8	62	2	S60788	M protein precurs
132	7	1.0	564	2	S76672	205	6	0.8	63	2	F69114	hypothetical prote
133	7	1.0	564	2	T00135	206	6	0.8	63	2	S44634	hypothetical prote
134	7	1.0	574	2	B29677	207	6	0.8	65	2	S97667	thiG protein AGR C
135	7	1.0	577	2	E72396	208	6	0.8	65	2	AF2891	thiamin biosynthes
136	7	1.0	582	2	T51625	209	6	0.8	65	4	IMBP14	hypothetical immun
137	7	1.0	583	2	B85063	210	6	0.8	67	2	G82702	hypothetical prote
138	7	1.0	585	2	C82955	211	6	0.8	67	2	H97845	hypothetical prote
139	7	1.0	608	2	G96575	212	6	0.8	67	2	A13396	hypothetical prote
140	7	1.0	632	2	T18692	213	6	0.8	68	2	S60797	M protein precurs
141	7	1.0	635	2	C18661	214	6	0.8	69	2	D75517	hypothetical prote
142	7	1.0	661	2	S19646	215	6	0.8	70	1	R5EC31	ribosomal protein
143	7	1.0	702	2	T23651	216	6	0.8	70	2	G86083	50S ribosomal subu
144	7	1.0	711	2	S88443	217	6	0.8	70	2	G12336	50S ribosomal subu
145	7	1.0	711	2	C70606	218	6	0.8	70	2	T41875	P10 orf137 - Bomby
146	7	1.0	714	2	S68993	219	6	0.8	73	2	E82344	rft-related prote
147	7	1.0	716	1	WZBEE6	220	6	0.8	75	2	S62073	M-like protein pre
148	7	1.0	731	2	A83410	221	6	0.8	76	2	S60810	M protein precurs
149	7	1.0	747	2	B47093	222	6	0.8	76	2	S61818	M-like protein ann
150	7	1.0	759	1	C64345	223	6	0.8	78	2	D34509	ig light chain C r
151	7	1.0	771	2	AF2056	224	6	0.8	78	2	A97974	degenerate transpo
152	7	1.0	778	2	G98221	225	6	0.8	79	2	S62072	protein M-like pro
153	7	1.0	778	2	A13064	226	6	0.8	79	2	S61802	M-like protein enn
154	7	1.0	778	2	T43223	227	6	0.8	79	2	S62078	M-like protein pre
155	7	1.0	778	2	T45221	228	6	0.8	80	2	S61801	M-like protein enn
156	7	1.0	809	2	S28999	229	6	0.8	81	2	G98177	M-like protein enn
157	7	1.0	811	2	T39998	230	6	0.8	81	2	G98177	hypothetical prote
158	7	1.0	819	2	S15169	231	6	0.8	82	2	S61816	M-like protein enn
159	7	1.0	823	2	S48911	232	6	0.8	82	2	B90706	hypothetical prote
160	7	1.0	863	2	S74447	233	6	0.8	82	2	S29722	cellular adhesion
161	7	1.0	869	2	A82055	234	6	0.8	82	2	B28262	hypothetical prote
162	7	1.0	875	2	T12794	235	6	0.8	82	2	E85556	hypothetical prote
163	7	1.0	960	1	A39651	236	6	0.8	82	2	B64791	ybdJ protein - Esc
164	7	1.0	963	2	T09478	237	6	0.8	83	2	S61086	M protein precurs
165	7	1.0	985	2	S67255	238	6	0.8	83	2	S61824	M-like protein enn
166	7	1.0	987	2	T40241	239	6	0.8	83	2	S62075	M-like protein enn
167	7	1.0	1029	2	F96310	240	6	0.8	84	2	C90447	partial transposas
168	7	1.0	1042	2	T13647	241	6	0.8	85	2	S62083	M-like protein enn
169	7	1.0	1057	2	H83273	242	6	0.8	85	2	S14026	hypothetical prote
170	7	1.0	1068	2	E81965	243	6	0.8	86	2	S62077	M-like protein enn
171	7	1.0	1082	2	H81020	244	6	0.8	87	2	T15407	hypothetical prote
172	7	1.0	1158	2	A83285	245	6	0.8	88	2	S60795	M protein precurs
173	7	1.0	1174	2	C97686	246	6	0.8	88	2	S60812	M protein precurs
174	7	1.0	1174	2	A82911	247	6	0.8	89	2	S61821	M-like protein enn
175	7	1.0	1239	2	I49705	248	6	0.8	90	2	S62074	M-like protein enn

249	6	0.8	90	2	S62081	M-like protein pre	322	6	0.8	131	2	A12329	30S ribosomal prot
250	6	0.8	90	2	B64031	hypothetical prote	323	6	0.8	131	2	C65089	hypothetical prote
251	6	0.8	90	2	D81684	conserved hypotet	324	6	0.8	132	2	AD0108	conserved hypotet
252	6	0.8	91	2	G90824	hypothetical prote	325	6	0.8	132	2	G72462	hypothetical prote
253	6	0.8	92	2	S61809	M-like protein em	326	6	0.8	132	2	A13314	acyl-CoA hydrolase
254	6	0.8	92	2	A84637	hypothetical prote	327	6	0.8	132	2	T50779	copper chaperone h
255	6	0.8	93	2	D69960	hypothetical prote	328	6	0.8	133	2	F64700	biopolymer transpo
256	6	0.8	94	2	S60858	M protein precuro	329	6	0.8	133	2	B82550	conserved hypotet
257	6	0.8	94	2	S61805	M-like protein em	330	6	0.8	133	2	T14238	probable secreted
258	6	0.8	94	2	S62084	M-like protein em	331	6	0.8	133	2	AE1971	hypothetical prote
259	6	0.8	94	2	B72867	fibrous body prote	332	6	0.8	134	2	T25527	hypothetical prote
260	6	0.8	96	2	D47624	Ig heavy chain V-I	333	6	0.8	134	2	B97643	hypothetical prote
261	6	0.8	96	2	S61827	M-like protein em	334	6	0.8	134	2	AE2866	hypothetical prote
262	6	0.8	96	2	S71519	M protein type 19	335	6	0.8	136	2	I49013	thymic shared anti
263	6	0.8	96	2	S10069	regulatory protein	336	6	0.8	137	2	P90565	30S ribosomal prot
264	6	0.8	97	2	S61811	M-like protein em	337	6	0.8	137	2	S34620	probable cytochrom
265	6	0.8	97	2	AF3595	hypothetical cytos	338	6	0.8	137	2	B75447	hypothetical prote
266	6	0.8	98	2	S61079	M protein precuro	339	6	0.8	137	2	A97256	uncharacterized pr
267	6	0.8	98	2	AB2100	hypothetical prote	340	6	0.8	138	1	R3NT11	ribosomal protein
268	6	0.8	99	2	D69060	hypothetical prote	341	6	0.8	138	1	R3PM11	ribosomal protein
269	6	0.8	101	2	C72502	hypothetical prote	342	6	0.8	138	1	QGBPL	host-nuclease inh
270	6	0.8	101	2	B44503	p12 protein - beet	343	6	0.8	138	2	B23525	ribosomal protein
271	6	0.8	102	2	B71503	probable thioredox	344	6	0.8	138	2	H90775	host-nuclease inh
272	6	0.8	102	2	B34509	Ig light chain C r	345	6	0.8	138	2	B83366	hypothetical prote
273	6	0.8	103	1	GERBM1	matrix Gla protein	346	6	0.8	139	1	RKQMSY	ribulose-bisphosph
274	6	0.8	103	2	AB4140	PTS system, cellob	347	6	0.8	139	2	S78312	Ig light chain C r
275	6	0.8	103	2	AB5821	unknown protein en	348	6	0.8	140	2	A34509	hypothetical prote
276	6	0.8	103	2	E90973	hypothetical prote	349	6	0.8	140	2	S74052	hypothetical prote
277	6	0.8	103	2	AD0713	histone H4 - Oxytr	350	6	0.8	141	2	E69260	hypothetical prote
278	6	0.8	104	2	J80154	histone H4 - Oxytr	351	6	0.8	141	2	D87402	hypothetical prote
279	6	0.8	104	2	S14184	histone H4 (clone	352	6	0.8	141	2	S08523	hypothetical prote
280	6	0.8	104	2	S55062	protein L-asparyl	353	6	0.8	141	2	S34617	hypothetical prote
281	6	0.8	104	2	E97300	hypothetical prote	354	6	0.8	143	1	FEYTA	ferredoxin [Fe-4S
282	6	0.8	105	2	S62082	M-like protein em	355	6	0.8	144	1	JC4266	interleukin-3 prec
283	6	0.8	105	2	T23564	hypothetical prote	356	6	0.8	144	2	S23658	superoxide dismuta
284	6	0.8	105	2	G69790	conserved hypotet	357	6	0.8	144	2	S32235	ribosomal protein
285	6	0.8	109	2	G75484	hypothetical prote	358	6	0.8	145	2	T51740	RNA helicase RH6 l
286	6	0.8	109	2	D81358	probable integral	359	6	0.8	145	2	S14185	histone H4 (clone
287	6	0.8	110	2	S70789	cagc protein precu	360	6	0.8	145	2	T10266	ribosomal protein
288	6	0.8	110	2	E90806	probable curli pro	361	6	0.8	145	2	H70946	hypothetical prote
289	6	0.8	110	2	A85666	conserved hypotet	362	6	0.8	146	1	I46407	interleukin-3 prec
290	6	0.8	111	2	C69168	conserved hypotet	363	6	0.8	146	2	A86490	protein T32520.11
291	6	0.8	111	2	G72745	hypothetical prote	364	6	0.8	146	2	G71181	hypothetical prote
292	6	0.8	113	2	T44302	hypothetical prote	365	6	0.8	147	2	G83301	conserved hypotet
293	6	0.8	113	2	D89762	hypothetical prote	366	6	0.8	148	1	RSRTLA	ribosomal protein
294	6	0.8	115	2	S44112	Ig heavy chain V r	367	6	0.8	148	1	RSMS27	ribosomal protein
295	6	0.8	116	2	S60818	M protein precuro	368	6	0.8	148	2	T51602	CAP binding protei
296	6	0.8	117	2	T12923	hypothetical prote	369	6	0.8	148	2	S25593	60S ribosomal prot
297	6	0.8	117	2	H84408	hypothetical prote	370	6	0.8	148	2	S60001	M-like protein em
298	6	0.8	118	2	G84322	hypothetical prote	371	6	0.8	148	2	C54128	hypothetical prote
299	6	0.8	119	2	E81272	hypothetical prote	372	6	0.8	148	2	T33937	ribosomal protein
300	6	0.8	120	2	T18059	hypothetical prote	373	6	0.8	149	1	R6NC7A	ribosomal protein
301	6	0.8	120	2	S61635	ribosomal protein	374	6	0.8	149	2	UC2392	ribosomal protein
302	6	0.8	121	2	D83624	probable two-compo	375	6	0.8	149	2	UC6202	ribosomal protein
303	6	0.8	122	1	S58649	ribosomal protein	376	6	0.8	149	2	D84292	transcription regu
304	6	0.8	122	2	AC1848	hypothetical prote	377	6	0.8	149	2	F82269	conserved hypotet
305	6	0.8	123	2	AG3530	hypothetical prote	378	6	0.8	149	2	B72694	hypothetical prote
306	6	0.8	126	2	G95154	conserved hypotet	379	6	0.8	151	1	MTBPA7	N-acetylmutamoyl-L
307	6	0.8	127	2	F81021	hypothetical prote	380	6	0.8	151	2	S07306	N-acetylmutamoyl-L
308	6	0.8	129	2	G69313	SSU ribosomal prot	381	6	0.8	151	2	A96493	hypothetical prote
309	6	0.8	129	2	S69854	hypothetical prote	382	6	0.8	151	2	T32957	hypothetical prote
310	6	0.8	129	2	T44125	hypothetical prote	383	6	0.8	151	2	S23215	hypothetical prote
311	6	0.8	129	2	G95303	hypothetical prote	384	6	0.8	151	2	E69077	hypothetical prote
312	6	0.8	129	2	AC3607	hypothetical prote	385	6	0.8	152	2	T44544	conserved hypotet
313	6	0.8	130	1	R3LV11	ribosomal protein	386	6	0.8	153	2	C98238	hypothetical. 16. 4k
314	6	0.8	130	2	S17719	H+-transporting tw	387	6	0.8	153	2	AH3047	conserved hypotet
315	6	0.8	130	2	S73215	ribosomal protein	388	6	0.8	153	2	H85638	hypothetical prote
316	6	0.8	130	2	S78255	ribosomal protein	389	6	0.8	154	2	T17124	hypothetical prote
317	6	0.8	130	2	T06945	ribosomal protein	390	6	0.8	154	2	D85093	hypothetical prote
318	6	0.8	130	2	T07518	ribosomal protein	391	6	0.8	154	2	A70508	hypothetical prote
319	6	0.8	130	2	C71023	hypothetical prote	392	6	0.8	154	2	C91171	probable beta-hydr
320	6	0.8	130	2	T43606	YopE chaperone b/c	393	6	0.8	154	2	C86017	probable beta-hydr
321	6	0.8	130	2	T31161	DNA primase homolo	394	6	0.8	155	2	T03463	hypothetical prote

395	6	0.8	156	1	R3TW7	ribosomal protein	468	6	0.8	182	2	S03259	methyl coenzyme M
396	6	0.8	156	1	JG0008	ribosomal protein	469	6	0.8	183	2	C75386	ribosome recycling
397	6	0.8	156	2	F69699	ribosomal protein	470	6	0.8	183	2	G64786	ybcL protein - Esc
398	6	0.8	156	2	T44379	ribosomal protein	471	6	0.8	183	2	T35733	hypothetical prote
399	6	0.8	156	2	B81234	30S ribosomal prot	472	6	0.8	184	2	T13186	hypothetical prote
400	6	0.8	156	2	AF1406	ribosomal protein	473	6	0.8	186	1	JQ1623	envelope-associate
401	6	0.8	156	2	AF1782	ribosomal protein	474	6	0.8	186	2	JQ1987	hypothetical 21.0K
402	6	0.8	156	2	AF2210	hypothetical prote	475	6	0.8	186	2	B83727	hypothetical prote
403	6	0.8	157	2	S23850	ribosomal protein	476	6	0.8	186	2	S51022	M21 chain - chicke
404	6	0.8	157	2	B87117	conserved hypot het	477	6	0.8	186	2	E97353	hypothetical prote
405	6	0.8	157	2	T24917	hypothetical prote	478	6	0.8	187	2	T00427	ribosomal protein
406	6	0.8	158	2	G82494	conserved hypot het	479	6	0.8	187	2	H84916	60S ribosomal prot
407	6	0.8	158	2	C32055	nirX protein - Syn	480	6	0.8	187	2	E90191	conserved hypot het
408	6	0.8	159	2	S75802	kdtB protein - Azo	481	6	0.8	187	2	A71328	probable survival
409	6	0.8	159	2	T48837	hypothetical prote	482	6	0.8	188	1	R5RT81	ribosomal protein
410	6	0.8	159	2	B83339	hypothetical prote	483	6	0.8	188	1	JN0779	ribosomal protein
411	6	0.8	160	2	A71723	ribosomal protein	484	6	0.8	188	1	R5XL6A	ribosomal protein
412	6	0.8	160	2	E97721	30S ribosomal prot	485	6	0.8	188	1	R5XL14	ribosomal protein
413	6	0.8	160	2	E95882	hypothetical trans	486	6	0.8	188	2	S83352	ribosomal protein
414	6	0.8	160	2	T16043	hypothetical prote	487	6	0.8	188	2	S83352	interferon alpha-1
415	6	0.8	161	2	T01651	hypoxically induce	488	6	0.8	189	2	IVMSA5	transcription regu
416	6	0.8	162	2	T11802	NADH2 dehydrogenas	489	6	0.8	189	2	H96903	interferon alpha-2
417	6	0.8	162	2	T45343	hypothetical prote	490	6	0.8	190	1	IVMSA2	interferon alpha-1
418	6	0.8	162	2	C72665	hypothetical prote	491	6	0.8	190	2	A24401	interferon alpha-1
419	6	0.8	163	2	F81374	hypothetical prote	492	6	0.8	190	2	F72206	hypothetical prote
420	6	0.8	163	2	T06903	ubiquinol-cytochro	493	6	0.8	191	2	F72731	hypothetical prote
421	6	0.8	164	2	JC2572	hypothetical 18K p	494	6	0.8	192	2	AD2530	hypothetical prote
422	6	0.8	164	2	S35021	nolB protein - Rhi	495	6	0.8	192	2	G70102	Holliday junction
423	6	0.8	165	2	T21313	hypothetical prote	496	6	0.8	197	2	AF2356	hypothetical prote
424	6	0.8	165	2	F97519	hypothetical prote	497	6	0.8	198	2	S37970	succinate dehydrog
425	6	0.8	165	2	AH2738	acetyltransferase	498	6	0.8	198	2	G82265	syd protein VC0903
426	6	0.8	166	2	E71342	probable deoxycyti	499	6	0.8	198	2	A72760	hypothetical prote
427	6	0.8	166	2	C90029	hypothetical prote	500	6	0.8	199	1	S47726	outer-membrane lip
428	6	0.8	167	1	JC1102	endothelin 3 precu	501	6	0.8	199	2	S35441	serine-tRNA ligase
429	6	0.8	167	2	B53293	flm3 region hypot h	502	6	0.8	199	2	A91176	outer membrane pro
430	6	0.8	167	2	S70824	hypothetical prote	503	6	0.8	199	2	B86022	outer membrane pro
431	6	0.8	167	2	F97238	probable acetylra	504	6	0.8	200	2	T05293	hypothetical prote
432	6	0.8	168	1	CYFGG2	gamma-crystallin I	505	6	0.8	201	2	T23192	hypothetical prote
433	6	0.8	168	2	C75425	hypothetical prote	506	6	0.8	201	2	T46404	hypothetical prote
434	6	0.8	169	2	T01783	gamma-2-crystallin	507	6	0.8	202	2	A82160	hypothetical prote
435	6	0.8	169	2	AB1010	formylmethionine d	508	6	0.8	202	2	B87342	hypothetical prote
436	6	0.8	169	2	C75427	hypothetical prote	509	6	0.8	203	2	I49153	hypothetical prote
437	6	0.8	170	2	S77192	general secretion	510	6	0.8	203	2	JC4845	cardiotrophin-1 -
438	6	0.8	170	2	F83325	hypothetical prote	511	6	0.8	203	2	T08332	conserved hypot het
439	6	0.8	170	2	C81967	probable acetylra	512	6	0.8	203	2	T48718	mitochondrial ribo
440	6	0.8	171	2	E87417	transcription regu	513	6	0.8	204	2	B90112	60S ribosomal prot
441	6	0.8	172	2	F97320	acetyltransferase	514	6	0.8	204	2	T19699	hypothetical prote
442	6	0.8	173	2	G86905	hypothetical prote	515	6	0.8	205	2	I40179	hypothetical prote
443	6	0.8	173	2	H82671	hypothetical prote	516	6	0.8	205	2	G83218	heme acquisition p
444	6	0.8	174	2	AF1668	hypothetical prote	517	6	0.8	206	2	F95113	phosphoglycerate m
445	6	0.8	174	2	AH1296	hypothetical prote	518	6	0.8	206	2	F84087	hypothetical prote
446	6	0.8	175	1	VGXRNH	glycoprotein NCVp5	519	6	0.8	207	2	E95024	ribosomal protein
447	6	0.8	175	1	VGXRE6	glycoprotein NCVp5	520	6	0.8	207	2	E97895	50S ribosomal prot
448	6	0.8	175	1	VGXRRB	glycoprotein NCVp5	521	6	0.8	207	2	F84203	transcription regu
449	6	0.8	175	1	VGXRRP5	glycoprotein NCVp5	522	6	0.8	207	2	F84518	hypothetical prote
450	6	0.8	175	2	T49053	hypothetical prote	523	6	0.8	208	2	AE0344	uracil phosphoribo
451	6	0.8	175	2	C97138	hypothetical prote	524	6	0.8	208	2	AH0818	uracil phosphoribo
452	6	0.8	176	2	C87070	conserved hypot het	525	6	0.8	208	2	AB3406	l-lactate permease
453	6	0.8	176	2	H70577	hypothetical prote	526	6	0.8	208	2	S44298	probable orotate p
454	6	0.8	177	2	T31242	hypothetical prote	527	6	0.8	209	2	H81651	conserved hypot het
455	6	0.8	177	2	S54778	NR-13 protein - qu	528	6	0.8	209	2	T19847	hypothetical prote
456	6	0.8	177	2	F95912	hypothetical prote	529	6	0.8	209	2	D83034	conserved hypot het
457	6	0.8	178	2	F82952	ATP synthase delta	530	6	0.8	209	2	JE0154	mitochondrial inne
458	6	0.8	178	2	S24298	chorion protein -	531	6	0.8	210	2	A71876	probable aldehyde
459	6	0.8	178	2	AB0870	outer membrane pro	532	6	0.8	210	2	B64639	oxygen-insensitive
460	6	0.8	179	2	T20206	hypothetical prote	533	6	0.8	210	2	I40540	vsd protein - Pse
461	6	0.8	180	2	T02388	hypothetical prote	534	6	0.8	210	2	F64609	conserved hypot het
462	6	0.8	181	1	TXSPM	thioredoxin m prec	535	6	0.8	210	2	S28673	hypothetical prote
463	6	0.8	181	2	T24914	hypothetical prote	536	6	0.8	210	2	E87005	probable thymidyla
464	6	0.8	181	2	AE0714	probable chorismat	537	6	0.8	210	2	T49785	hypothetical prote
465	6	0.8	181	2	F82750	outer membrane pro	538	6	0.8	211	2	F83193	L-isoeaspartate pro
466	6	0.8	182	2	B53311	acetyl-CoA carboxy	539	6	0.8	211	2	H82803	conserved hypot het
467	6	0.8	182	2	A12437	biotin carboxyl ca	540	6	0.8	211	2	AB2026	transcription term

541	6	0.8	213	2	S01766	GTP-binding protei	614	6	0.8	238	1	UC2297	CD63 antigen - rab
542	6	0.8	213	2	E34323	GTP-binding protei	615	6	0.8	238	1	S43511	CD63/ME491 antigen
543	6	0.8	213	2	G75521	ABC transporter, A	616	6	0.8	238	1	A46508	CD63/ME491 antigen
544	6	0.8	213	2	S43723	lactase (EC 3.2.1.	617	6	0.8	238	2	T10771	NADPH-ferrithemop
545	6	0.8	213	2	AF0002	molycoplerin-guan	618	6	0.8	238	2	S70468	NADPH-ferrithemop
546	6	0.8	214	2	T05176	hypothetical prote	619	6	0.8	238	2	D83847	agglutinin (WBA I)
547	6	0.8	214	2	S27523	hypothetical prote	620	6	0.8	238	2	D83847	two-component resp
548	6	0.8	215	2	T22572	hypothetical prote	621	6	0.8	238	2	A72374	hypothetical prote
549	6	0.8	215	2	S75345	hypothetical prote	622	6	0.8	238	2	H95105	pilin gene inverti
550	6	0.8	215	2	AD2407	hypothetical prote	623	6	0.8	239	2	F84189	hypothetical prote
551	6	0.8	216	2	H87547	heterocyst diffe	624	6	0.8	239	2	D96587	hypothetical prote
552	6	0.8	217	1	A65026	procalcitonin 3,	625	6	0.8	239	2	T15222	hypothetical prote
553	6	0.8	217	1	W0H02C	uracil phosphoribo	626	6	0.8	240	2	A72623	hypothetical prote
554	6	0.8	217	2	H91048	homeotic protein H	627	6	0.8	240	2	A41797	Ig light chain - s
555	6	0.8	217	2	D85893	uracil phosphoribo	628	6	0.8	240	2	A10941	probable GntR-fam
556	6	0.8	217	2	T31875	hypothetical prote	629	6	0.8	241	2	AC2646	conserved hypotet
557	6	0.8	218	2	F86844	hypothetical prote	630	6	0.8	241	2	AC1918	hypothetical prote
558	6	0.8	218	2	B97428	hypothetical prote	631	6	0.8	241	2	S74620	hypothetical prote
559	6	0.8	218	2	A34445	25K calcium-bindin	632	6	0.8	242	2	A42299	hypothetical prote
560	6	0.8	218	2	G70654	hypothetical prote	633	6	0.8	242	2	D81161	conserved hypotet
561	6	0.8	219	2	A44503	p25 protein - beet	634	6	0.8	242	2	H11806	probable outer mem
562	6	0.8	219	2	G64009	hypothetical prote	635	6	0.8	244	2	A55541	heterocyst maturat
563	6	0.8	220	2	T44702	hypothetical prote	636	6	0.8	244	2	A12269	heterocyst specifi
564	6	0.8	220	2	S75751	hypothetical prote	637	6	0.8	244	2	A69180	precocorin-3 methyl
565	6	0.8	221	2	S20964	ribosomal protein	638	6	0.8	245	1	A41021	NAD+-asparagine AD
566	6	0.8	221	2	G84169	hypothetical prote	639	6	0.8	245	1	C65210	hypothetical 26..3
567	6	0.8	221	2	S54452	hypothetical prote	640	6	0.8	245	2	A72218	guFA protein - The
568	6	0.8	222	2	S26994	cellulose 1,4-beta	641	6	0.8	245	2	C91255	hypothetical prote
569	6	0.8	222	2	S20029	homeotic protein H	642	6	0.8	245	2	G86095	hypothetical prote
570	6	0.8	222	2	A81944	hypothetical prote	643	6	0.8	245	2	B95943	probable choline u
571	6	0.8	222	2	I39192	gene HOXA1 protein	644	6	0.8	246	2	186673	gene X123 protein
572	6	0.8	222	2	G84175	regulatory protein	645	6	0.8	246	2	S27727	hypothetical prote
573	6	0.8	223	1	A41965	transcription regu	646	6	0.8	246	2	A10920	probable UDP-N-ace
574	6	0.8	223	2	A81285	probable triose-ph	647	6	0.8	246	2	AH3072	transcription regu
575	6	0.8	223	2	B90829	transcription regu	648	6	0.8	248	1	J01692	infected cell prot
576	6	0.8	223	2	H85686	transcription regu	649	6	0.8	248	2	AC1657	glucose 1-dehydrog
577	6	0.8	223	2	AB0199	response regulator	650	6	0.8	248	2	AC1285	glucose 1-dehydrog
578	6	0.8	224	1	RGBBFT	transcription regu	651	6	0.8	248	2	AH1274	probable cyclase C
579	6	0.8	224	2	AH0646	transcription regu	652	6	0.8	248	2	F72087	conserved hypotet
580	6	0.8	224	2	T10660	phosphatase II pro	653	6	0.8	248	2	D66537	CT056 hypothetical
581	6	0.8	224	2	T33606	hypothetical prote	654	6	0.8	248	2	B64703	outer membrane pro
582	6	0.8	225	2	G83371	probable amino aci	655	6	0.8	248	2	E71816	probable outer mem
583	6	0.8	225	2	T02660	germin-like protei	656	6	0.8	249	2	A82796	hypothetical prote
584	6	0.8	225	2	T04210	hypothetical prote	657	6	0.8	249	2	T05424	hypothetical prote
585	6	0.8	225	2	F89795	hypothetical prote	658	6	0.8	250	2	H82079	protein disulfide-
586	6	0.8	226	2	S70467	agglutinin (WBA I)	659	6	0.8	250	2	T10663	hypothetical prote
587	6	0.8	226	2	JC5327	adhesin complex 25	660	6	0.8	250	2	T35927	conserved hypotet
588	6	0.8	227	2	S01260	M protein precurs	661	6	0.8	250	2	A46957	NAD+-asparagine AD
589	6	0.8	228	2	S67199	hypothetical prote	662	6	0.8	251	2	B64102	ribosomal protein
590	6	0.8	229	2	H75205	hypothetical prote	663	6	0.8	251	2	G81279	enterocheilin uptak
591	6	0.8	230	2	T06709	protein-L-isoospar	664	6	0.8	251	2	F72114	conserved hypotet
592	6	0.8	231	2	S37108	cuticlin 2 - Caen	665	6	0.8	251	2	A86508	ACR family [import
593	6	0.8	231	2	AE1907	hypothetical prote	666	6	0.8	251	2	T32308	hypothetical prote
594	6	0.8	231	2	T44086	transcription regu	667	6	0.8	251	2	B64048	hypothetical prote
595	6	0.8	231	2	T02765	glutathione trans	668	6	0.8	252	2	S69786	pikK protein - Sal
596	6	0.8	231	2	C86665	amino acid ABC tra	669	6	0.8	252	2	AC0849	pathogenicity 1 is
597	6	0.8	231	2	B72100	hypothetical prote	670	6	0.8	252	2	D86462	ATG1-like protein,
598	6	0.8	231	2	B66524	hypothetical prote	671	6	0.8	252	2	B84205	hypothetical prote
599	6	0.8	232	1	A54361	venombin A (EC 3.4	672	6	0.8	253	2	T41139	proteasome compone
600	6	0.8	232	2	A87083	conserved hypotet	673	6	0.8	253	2	I69732	stationary-phase s
601	6	0.8	233	2	I51383	Ig lambda chain -	674	6	0.8	253	2	F91078	survival protein f
602	6	0.8	233	2	A10109	probable endonuc	675	6	0.8	253	2	G85923	stationary-phase s
603	6	0.8	233	2	T35251	probable endonuc	676	6	0.8	253	2	AB0856	stationary-phase s
604	6	0.8	234	2	C70156	hypothetical prote	677	6	0.8	253	2	H70780	hypothetical prote
605	6	0.8	236	2	T27514	hypothetical prote	678	6	0.8	253	2	AG2627	hypothetical prote
606	6	0.8	236	2	B56281	Talpa-cephe-meth	679	6	0.8	254	2	UC5004	hypothetical prote
607	6	0.8	236	2	C95387	hypothetical prote	680	6	0.8	254	2	AH0407	attacin A precurs
608	6	0.8	236	2	D72556	hypothetical prote	681	6	0.8	254	2	T35366	probable membrane
609	6	0.8	237	2	B69838	transcription regu	682	6	0.8	255	1	G69495	probable membrane
610	6	0.8	237	2	C95199	nitroductase fam	683	6	0.8	255	1	A28169	venombin A (EC 3.4
611	6	0.8	237	2	A98066	NADPH-flavin oxido	684	6	0.8	255	2	S64574	probable membrane
612	6	0.8	237	2	B64643	hypothetical prote	685	6	0.8	255	2	G71007	thiamin biosynthes
613	6	0.8	238	1	I38016	melanoma-associate	686	6	0.8	255	2	A82671	conserved hypotet

687 6 0.8 256 2 F97555 triosephosphate is
688 6 0.8 256 2 AH2775 triosephosphate is
689 6 0.8 256 2 T12999 aquaporin homolog
690 6 0.8 256 2 S11877 chlorophyll a/b-bi
691 6 0.8 257 2 G95252 PTS system, IIC co
692 6 0.8 257 2 E98117 hypothetical prote
693 6 0.8 257 2 H84713 hypothetical prote
694 6 0.8 257 2 F75476 probable uroporphyr
695 6 0.8 257 2 T23332 hypothetical prote
696 6 0.8 258 2 D70341 hypothetical prote
697 6 0.8 258 2 S61656 hypothetical prote
698 6 0.8 258 2 S73803 MG256 homolog H91_
699 6 0.8 258 2 B97540 transmembrane regu
700 6 0.8 258 2 AD2759 transcription regu
701 6 0.8 258 2 H72202 ROK family protein
702 6 0.8 259 2 T34637 probable type IV p
703 6 0.8 259 2 S41890 hypothetical prote
704 6 0.8 260 2 S22850 ERs1 protein yea
705 6 0.8 261 2 C87204 thioesterase limpo
706 6 0.8 261 2 AC3646 hydroxypyruvate is
707 6 0.8 261 2 T24532 hypothetical prote
708 6 0.8 261 2 E75626 hypothetical prote
709 6 0.8 261 2 S44776 C30A5 protein - Ca
710 6 0.8 263 2 D71281 probable pyrroline
711 6 0.8 263 2 F92311 survival protein S
712 6 0.8 263 2 S64472 hypothetical prote
713 6 0.8 264 2 AD4957 methionyl aminopep
714 6 0.8 264 2 AD3353 probable membrane
715 6 0.8 264 2 F70969 hypothetical prote
716 6 0.8 264 2 D97646 probable ABC-trans
717 6 0.8 264 2 AD2870 hypothetical prote
718 6 0.8 265 2 AF2007 hypothetical prote
719 6 0.8 266 1 ENSAC1 enterotoxin C-1 pr
720 6 0.8 266 2 E69083 thiamin biosynthes
721 6 0.8 266 2 C95288 probable ABC trans
722 6 0.8 267 1 A64933 metastasis suppres
723 6 0.8 267 2 B72408 conserved hypoteth
724 6 0.8 267 2 T15595 hypothetical prote
725 6 0.8 267 2 C72415 zinc ABC transport
726 6 0.8 267 2 H97299 proB [imported] -
727 6 0.8 268 2 AG2594 enoyl-(acyl)-carrie
728 6 0.8 268 2 S68826 pancreatic elastas
729 6 0.8 268 2 S68825 pancreatic elastas
730 6 0.8 268 2 S67583 KXN1 protein - yea
731 6 0.8 268 2 S72968 cytotoxin/hemolysi
732 6 0.8 268 2 E70502 probable tlyA prot
733 6 0.8 268 2 B75279 survival protein S
734 6 0.8 268 2 T35568 probable lipoprote
735 6 0.8 268 2 B98223 branched-chain ami
736 6 0.8 268 2 AF3063 hypothetical prote
737 6 0.8 269 2 B44499 major intrinsic pr
738 6 0.8 269 2 A41616 erythrocyte integr
739 6 0.8 269 2 JCL320 water channel prot
740 6 0.8 269 2 I52366 uterine water chan
741 6 0.8 269 2 H87078 cytotoxin/hemolysi
742 6 0.8 269 2 C91030 probable structura
743 6 0.8 269 2 D85874 probable structura
744 6 0.8 269 2 AD0804 probable membrane
745 6 0.8 269 2 E98165 hypothetical prote
746 6 0.8 269 2 AB3122 hypothetical prote
747 6 0.8 269 2 S08346 probable structura
748 6 0.8 270 2 H97376 enoyl-(acyl)-carrie
749 6 0.8 270 2 AF1172 mannose-specific p
750 6 0.8 270 2 F72381 D-tagatose 3-epime
751 6 0.8 271 2 AG1529 mannose-specific p
752 6 0.8 271 2 BC4861 50S ribosomal prot
753 6 0.8 271 2 J62348 water channel prot
754 6 0.8 271 2 S75967 hypothetical prote
755 6 0.8 271 2 T21777 hypothetical prote
756 6 0.8 271 2 AF2234 glutamine-binding
757 6 0.8 272 2 AH3496 enoyl-[acyl]-carrie
758 6 0.8 272 2 C72498 probable stress pr
759 6 0.8 272 2 B75293 amino acid ABC tra

760 6 0.8 272 2 E70862 hypothetical prote
761 6 0.8 272 2 H87395 hypothetical prote
762 6 0.8 272 2 C96742 unknown protein F1
763 6 0.8 272 2 H71124 hypothetical prote
764 6 0.8 272 2 A72512 probable hypoxanth
765 6 0.8 273 2 G89856 conserved hypoteth
766 6 0.8 273 2 T08645 yfjQ protein - Esc
767 6 0.8 273 2 T40611 hypothetical prote
768 6 0.8 273 2 S21731 signal recognition
769 6 0.8 274 2 D95165 NH(3)-dependent NA
770 6 0.8 274 2 C98031 NAD synthase (glut
771 6 0.8 274 2 T06866 nosineptide resist
772 6 0.8 274 2 T20435 hypothetical prote
773 6 0.8 274 2 E83909 oligopeptide ABC t
774 6 0.8 275 2 T17966 ribonuclease III h
775 6 0.8 275 2 F70191 rare lipoprotein A
776 6 0.8 275 2 S76233 transcription term
777 6 0.8 276 2 D41044 octopine-binding p
778 6 0.8 276 2 G75358 hypothetical prote
779 6 0.8 276 2 T29689 hypothetical prote
780 6 0.8 276 2 D90595 hypothetical prote
781 6 0.8 277 2 B83353 probable binding-p
782 6 0.8 277 2 B87651 conserved hypoteth
783 6 0.8 278 2 D64750 yafZ protein - Esc
784 6 0.8 278 2 G75177 dipeptide transpor
785 6 0.8 278 2 S09662 hypothetical prote
786 6 0.8 278 2 A13633 hypothetical prote
787 6 0.8 279 2 C70458 diaminomelate ep
788 6 0.8 279 2 AG3603 maltose transport
789 6 0.8 279 2 C87444 conserved hypoteth
790 6 0.8 279 2 E75573 transcription regu
791 6 0.8 280 2 G72428 oligopeptide ABC t
792 6 0.8 280 2 S64246 hypothetical prote
793 6 0.8 280 2 S48903 SOL3 protein - yea
794 6 0.8 280 2 T30080 hypothetical prote
795 6 0.8 281 2 AC3471 cysQ protein limpo
796 6 0.8 281 2 T22280 hypothetical prote
797 6 0.8 281 2 T39199 C2H2-150 - human
798 6 0.8 282 2 JS0168 collagen col-8 - C
799 6 0.8 282 2 T16036 cuticle collagen c
800 6 0.8 283 2 A64174 hypothetical prote
801 6 0.8 284 2 T52062 PRAG-like protein
802 6 0.8 285 2 E75317 glycerophosphoryl
803 6 0.8 285 2 A97561 hypothetical prote
804 6 0.8 286 2 H64054 site-specific DNA-
805 6 0.8 286 2 H88175 protein T24H7.1 [i
806 6 0.8 286 2 T34096 hypothetical prote
807 6 0.8 286 2 B95011 hypothetical prote
808 6 0.8 287 2 T22605 hypothetical prote
809 6 0.8 288 2 H84757 hypothetical prote
810 6 0.8 289 2 H95929 probable hydroxyme
811 6 0.8 289 2 T02839 probable membrane
812 6 0.8 289 2 A99367 hypothetical prote
813 6 0.8 289 2 E81259 hypothetical prote
814 6 0.8 290 2 JX0175 lectin precursor -
815 6 0.8 290 2 T09630 transcription regu
816 6 0.8 290 2 S53904 hypothetical prote
817 6 0.8 290 2 C90382 membrane conserved
818 6 0.8 290 2 H64431 glycosyl transfera
819 6 0.8 291 2 T04122 preprotein translo
820 6 0.8 291 2 T11983 thiamin biosynthes
821 6 0.8 291 2 E95106 carbon-nitrogen hy
822 6 0.8 291 2 G97974 beta-ureidopropion
823 6 0.8 291 2 B90256 conserved hypoteth
824 6 0.8 291 2 A72341 hypothetical prote
825 6 0.8 291 2 G95139 geranyltransf
826 6 0.8 291 2 E84646 hypothetical prote
827 6 0.8 292 2 F98007 geranyltransf
828 6 0.8 292 1 B35124 biphenyl-2,3-diol
829 6 0.8 292 2 G75399 N-acetylglutamate
830 6 0.8 292 2 T09531 2-dehydro-3-deoxy-
831 6 0.8 293 2 AC2215 UTP-glucose-1-phos
832 6 0.8 293 2 A89818 hypothetical prote

833	6	0.8	294	2	B87210	hypothetical prote	906	6	0.8	311	2	B90715	probable tRNA synt
834	6	0.8	294	2	AB0115	probable carbon-ni	907	6	0.8	311	2	B85565	probable tRNA synt
835	6	0.8	294	2	B86870	hypothetical prote	908	6	0.8	311	2	AB0583	probable nucleosid
836	6	0.8	294	2	F64337	hypothetical prote	909	6	0.8	312	2	C70475	cytochrome c bioge
837	6	0.8	294	2	B69759	hypothetical prote	910	6	0.8	312	2	G71146	hypothetical prote
838	6	0.8	294	2	D87505	geranyltransstranf	911	6	0.8	312	2	G84973	tRNA pseudouridine
839	6	0.8	294	2	G95861	probable inositol	912	6	0.8	312	2	A42791	conserved hypotnet
840	6	0.8	295	2	AB2156	hypothetical prote	913	6	0.8	312	2	T20932	hypothetical prote
841	6	0.8	296	2	S72943	diaminopimelate ep	914	6	0.8	313	2	B69588	L-arabinose transp
842	6	0.8	296	2	T48926	bzip transcription	915	6	0.8	313	2	B64896	probable membrane
843	6	0.8	296	2	T18993	microtubule-associ	916	6	0.8	313	2	A85734	probable transport
844	6	0.8	296	2	T27768	hypothetical prote	917	6	0.8	313	2	P80884	probable transport
845	6	0.8	296	2	H81726	UDP-N-acetylglucos	918	6	0.8	313	2	S60713	polygalacturonase-
846	6	0.8	297	1	ROBO	thiosulfate sulfur	919	6	0.8	313	2	S30954	minor tail protein
847	6	0.8	297	1	JC5286	geranyltransstranf	920	6	0.8	313	2	D75403	hypothetical prote
848	6	0.8	298	1	JN0257	biphenyl-2,3-diol	921	6	0.8	313	2	A48903	beta-lactamase - P
849	6	0.8	298	1	JA2409	biphenyl-2,3-diol	922	6	0.8	314	2	G95403	probable ABC trans
850	6	0.8	298	1	T20841	hypothetical prote	923	6	0.8	314	2	H70840	probable regulator
851	6	0.8	298	2	T20841	hypothetical prote	924	6	0.8	314	2	H70801	hypothetical prote
852	6	0.8	299	2	T01806	hypothetical prote	925	6	0.8	314	2	H89849	conserved hypotnet
853	6	0.8	299	2	C90057	hypothetical prote	926	6	0.8	314	2	AE3193	conserved hypotnet
854	6	0.8	300	2	A71645	protein p34 (p34)	927	6	0.8	314	2	G83487	hypothetical prote
855	6	0.8	300	2	S72756	cytochrome O ubiqn	928	6	0.8	314	2	A70503	hypothetical prote
856	6	0.8	300	2	B69053	hypothetical prote	929	6	0.8	315	2	H97570	hypothetical 34.4K
857	6	0.8	300	2	T00274	hypothetical prote	930	6	0.8	315	2	I50706	transcription fact
858	6	0.8	300	2	H86358	zinc finger protei	931	6	0.8	316	2	S50336	NADH2 dehydrogenas
859	6	0.8	301	1	RDHUB5	cytochrome-b5 redu	932	6	0.8	316	2	B90282	hypothetical prote
860	6	0.8	301	2	H95308	probable ABC trans	933	6	0.8	316	2	D83406	probable transmemb
861	6	0.8	301	2	S72642	protein R11G11.15	934	6	0.8	316	2	T08694	hypothetical prote
862	6	0.8	301	2	A88931	hypothetical prote	935	6	0.8	316	2	T28072	hypothetical prote
863	6	0.8	301	2	A98214	hypothetical prote	936	6	0.8	316	2	G82696	rod shape-determin
864	6	0.8	302	2	T34398	hypothetical prote	937	6	0.8	317	2	F71139	hypothetical prote
865	6	0.8	302	2	T03066	probable palmitoyl	938	6	0.8	317	2	E75421	conserved hypotnet
866	6	0.8	303	1	B36227	urate oxidase (EC	939	6	0.8	317	2	G83033	vanillate O-demeth
867	6	0.8	303	2	A45640	phosphoprotein pho	940	6	0.8	317	2	F89349	D-ribose-binding p
868	6	0.8	303	2	S74949	hypothetical prote	941	6	0.8	317	2	A12932	hypothetical prote
869	6	0.8	303	2	B84296	hypothetical prote	942	6	0.8	317	2	AE1932	hypothetical prote
870	6	0.8	303	2	A71819	hypothetical prote	943	6	0.8	318	2	C82231	chloroedoxin reduct
871	6	0.8	303	2	F64701	hypothetical prote	944	6	0.8	318	2	B72676	hypothetical prote
872	6	0.8	303	2	H84758	hypothetical prote	945	6	0.8	318	2	A99510	hypothetical prote
873	6	0.8	303	2	F84401	hypothetical prote	946	6	0.8	318	2	AD3200	conserved hypotnet
874	6	0.8	303	2	S74517	hypothetical prote	947	6	0.8	319	2	S61421	aldose reductase h
875	6	0.8	303	2	C97463	hypothetical prote	948	6	0.8	319	2	C87447	sulphate ABC trans
876	6	0.8	304	1	WZBEA9	33.2K tegument pro	949	6	0.8	320	1	S15024	aldose reductase-x
877	6	0.8	304	2	B82089	transcription regu	950	6	0.8	320	2	J02253	aldehyde reductase
878	6	0.8	304	2	AG1182	hypothetical prote	951	6	0.8	320	2	T48188	aldose reductase-1
879	6	0.8	304	2	AG0759	pdu/cob regulatory	952	6	0.8	320	2	B70947	probable moxR3 pro
880	6	0.8	305	2	F72574	probable cytochrom	953	6	0.8	320	2	AH0067	conserved hypotnet
881	6	0.8	305	2	H75061	hypothetical prote	954	6	0.8	320	2	F75348	hypothetical prote
882	6	0.8	306	2	S19997	hypothetical prote	955	6	0.8	320	2	S61586	probable membrane
883	6	0.8	306	2	T32801	hypothetical prote	956	6	0.8	321	2	H86981	probable prochrome
884	6	0.8	307	2	D87688	hypothetical prote	957	6	0.8	321	2	I18238	transcription fact
885	6	0.8	307	2	S51373	tau-protein kinase	958	6	0.8	322	2	UN0419	DNA-directed RNA p
886	6	0.8	307	2	S50850	cyclin-dependent k	959	6	0.8	322	2	F83140	thiamine-phosphate
887	6	0.8	307	2	S50861	cyclin-dependent k	960	6	0.8	322	2	T48460	MADS-box protein-1
888	6	0.8	307	2	C83188	probable ATP-bindi	961	6	0.8	323	2	T01626	peroxidase (EC 1.1
889	6	0.8	308	2	B82394	hypothetical prote	962	6	0.8	323	2	A72508	probable cobalamin
890	6	0.8	308	2	T08796	tropomyosin - huma	963	6	0.8	323	2	F84423	hypothetical prote
891	6	0.8	308	2	C83586	probable transcrip	964	6	0.8	323	2	H90119	hypothetical prote
892	6	0.8	308	2	T14739	hypothetical prote	965	6	0.8	324	1	B69553	methanol dehydroge
893	6	0.8	309	2	D75008	methanol dehydroge	966	6	0.8	324	2	C69261	2-hydroxyhepta-2,4
894	6	0.8	309	2	F72316	hypothetical prote	967	6	0.8	324	2	T28032	hypothetical prote
895	6	0.8	309	2	B38545	hypothetical prote	968	6	0.8	324	2	D95218	iron-compound ABC
896	6	0.8	309	2	F71824	hypothetical prote	969	6	0.8	324	2	B98082	hypothetical prote
897	6	0.8	310	2	A11059	carbamate kinase (970	6	0.8	324	2	G81330	probable phosphata
898	6	0.8	310	2	B87242	probable LysR faml	971	6	0.8	325	2	AE82073	GTP-binding protei
899	6	0.8	310	2	JC4343	uridine phosphoryl	972	6	0.8	325	2	A13635	thiamine biosynthe
900	6	0.8	310	2	H82439	hypothetical prote	973	6	0.8	325	2	AE3575	transcription regu
901	6	0.8	310	2	T29731	hypothetical prote	974	6	0.8	325	2	C91138	hypothetical prote
902	6	0.8	310	2	D71679	UDP-n-acetylglucos	975	6	0.8	325	2	F85983	hypothetical prote
903	6	0.8	311	2	T50326	probable mitochond	976	6	0.8	325	2	A10904	probable membrane
904	6	0.8	311	2	D64240	methionyl-tRNA for	977	6	0.8	325	2	F55110	hypothetical 34.7
905	6	0.8	311	2	A64800	purine nucleosidas	978	6	0.8	326	1	J01442	glycoprotein VP7 p

979 6 0.8 326 2 S14266 uracil-DNA glycosy
980 6 0.8 326 2 T45226 probable NS, N10-me
981 6 0.8 326 2 AG1316 protein-tyrosine/s
982 6 0.8 326 2 AG1688 protein-tyrosine/s
983 6 0.8 327 2 T10263 probable polygalac
984 6 0.8 327 2 B71146 probable glucose-1
985 6 0.8 327 2 D87645 sensor histidine k
986 6 0.8 328 2 A87410 myo-inositol 2-deh
987 6 0.8 329 2 T08262 hypothetical protei
988 6 0.8 329 2 S53494 RNA-binding protei
989 6 0.8 329 2 T00873 hypothetical prote
990 6 0.8 329 2 T20546 hypothetical prote
991 6 0.8 330 2 G75007 lps biosynthesis r
992 6 0.8 330 2 H64077 aspartate-ammonia
993 6 0.8 330 2 D95230 aspartate-ammonia
994 6 0.8 330 2 S40853 probable 2-keto-3-
995 6 0.8 330 2 C86080 2-keto-3-deoxy-D-g
996 6 0.8 330 2 C91233 2-keto-3-deoxy-D-g
997 6 0.8 330 2 S35439 transcription fact
998 6 0.8 330 2 AD1533 hypothetical prote
999 6 0.8 331 1 A70032 conserved hypothet
1000 6 0.8 331 2 F75074 hypothetical prote

ALIGNMENTS

RESULT 1
TonB-dependent receptor NMB1829 [imported] - Neisseria meningitidis (strain MC58 serog
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: B81038
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
xi H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Piza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: B81038
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <TET>
A:Cross-references: GB:AE002532; GB:AE002098; NID:g7227078; PIDN:AAF42164.1; PID:g722708
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1829

Query Match 23.8%; Score 172; DB 2; Length 708;
Best Local Similarity 100.0%; Pred. No. 6e-173;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 61 TVGTKIPASLREIPQSVSIITNQVKQRNVDTFDQLARKTPGLRVLSNDDGRSSVYARGY 120
Db 61 TVGTKIPASLREIPQSVSIITNQVKQRNVDTFDQLARKTPGLRVLSNDDGRSSVYARGY 120
Qy 121 EYSEYNIDGLPAQMOSINGTLPNLFADFVEVMRGPGLFDSGEMGGIVNLVRKRPTKA 180
Db 121 EYSEYNIDGLPAQMOSINGTLPNLFADFVEVMRGPGLFDSGEMGGIVNLVRKRPTKA 180
Qy 181 FQGHAAAGFGTHQYKAEADVSGSLNSDGSVRGVMAQTVGASPRPAEKNNR 232
Db 181 FQGHAAAGFGTHQYKAEADVSGSLNSDGSVRGVMAQTVGASPRPAEKNNR 232

RESULT 2
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
AC2079
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2079

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073884.1; PID:g17131276; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2185
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 1.5%; Score 11; DB 2; Length 853;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 627 PKHSANLWTTY 637
Db 751 PKHSANLWTTY 761

RESULT 3
C83997
1-deoxyxylulose-5-phosphate synthase BH2779 [imported] - Bacillus halodurans (strain C-1
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C83997
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C83997
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-629 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06498.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2779
C:Superfamily: hypothetical protein C2814

Query Match 1.2%; Score 9; DB 2; Length 629;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 470 GHYKIESGE 478
Db 299 GHYKIESGE 307

RESULT 4
D81976
probable ferric siderophore receptor protein NMA0575 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D81976
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: D81976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83866.1; PID:g737931
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA0575
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homo

Query Match 1.2%; Score 9; DB 2; Length 725;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LREIPOSVS 78
 |||||
 DB 68 LREIPOSVS 76

RESULT 5

H81030
 TonB-dependent receptor NMB1882 [imported] - Neisseria meningitidis (strain MC58 serogroup C)
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: H81030

R:Retelien, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A. Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; et al., H.; Qin, H.; Vamathevan, J.; Gill, U.; Scarlato, V.; Maignani, V.; Pizzi, M. Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; PMID:20175755; PMID:10710307

A:Accession: H81030
 A:Status: Preliminary
 A:Molecule type: DNA

A:Residues: 1-725 <TE>
 A:Cross-references: GB:AE002538; GB:AE002098; NID:g7227136; PIDN:AAFA2216.1; PID:g722714
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:

A:Gene: NMB1882
 C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom

Query Match 1.2%; Score 9; DB 2; Length 725;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 LREIPOSVS 78
 |||||
 DB 68 LREIPOSVS 76

RESULT 6

TJ39813
 Hypothetical protein SPBC19C7.07c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: TJ39813
 R:Lyne, M.; Rajandream, M.A.; Bartell, B.G.; Lucas, M.; Gaillardin, C.

A:Reference number: 221881

A:Accession: TJ39813
 A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-284 <LYN>
 A:Cross-references: EMBL:AL023859; PIDN:CA119575.1; GSPDB:GN00067; SPDB:SPBC19C7.07c

A:Experimental source: strain 972h-, cosmid c19C7

C:Genetics:
 A:Gene: SPDB:SPBC19C7.07c

A:Map position: 2

Query Match 1.1%; Score 8; DB 2; Length 284;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 LIGGRLG 470
 |||||
 DB 233 LIGGRLG 240

RESULT 7

WBBF5
 59k transcription activator - equine herpesvirus 1 (strain Abap)

C:Species: equine herpesvirus 1
 A:Note: host Equus caballus (domestic horse)
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 17-Mar-2000
 C:Accession: J36801

R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
 submitted to GenBank, March 1992

A:Description: The DNA sequence of equine herpesvirus-1.

A:Reference number: A36805

A:Accession: J36801

A:Molecule type: DNA

A:Residues: 1-532 <TEL>
 A:Cross-references: GB:M6664; NID:g330791; PIDN:AB02498.1; PID:g330854

R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
 Virology 189, 304-316, 1992

A:Title: The DNA sequence of equine herpesvirus-1.

A:Reference number: A41831; PMID:92295566; PMID:1318606

A:Contents: annotation; possible protein-coding frames

A:Note: neither amino acid nor nucleotide sequence is given

C:Genetics:
 A:Gene: 63

C:Superfamily: varicella-zoster virus gene 63 protein; RING finger homology

C:Keywords: DNA binding; transcription regulation; zinc finger

F:4-52/Domain: RING finger homology <RNG>
 F:8-46/Region: zinc finger C3HC4 motif

Query Match 1.1%; Score 8; DB 1; Length 532;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ADVSGSLN 206
 |||||
 DB 323 ADVSGSLN 330

RESULT 8
 G70794
 Probable alpha-isopropyl malate synthase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: G70794

R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Bartell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:9829587; PMID:9634230

A:Accession: G70794

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Cross-references: GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CA18032.1; PID:g296013

A:Experimental source: strain H37RV

C:Genetics:
 A:Gene: leuA

C:Superfamily: 2-isopropylmalate synthase

Query Match 1.1%; Score 8; DB 2; Length 644;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 585 GVEETISG 592
 |||||
 DB 522 GVEETISG 529

RESULT 9

AC0547
 ferrioxamine B receptor precursor [imported] - Salmonella enterica subsp. enterica serov. typhi
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AC0547

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, Th, T.; Connerston, P.; Cronin, A.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulle, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AC0547
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-696 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD08819.1; PID:gl6501633; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY0396

Query Match 1.1%; Score 8; DB 2; Length 696;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 PQSVSIIT 81
 |||||
 Db 54 PQSVSIIT 61

RESULT 10

A36942
 Fe(III)-pyochelin receptor fptA precursor - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 31-Dec-2000
 C:Accession: A36942; G83117
 R;Ankenbauer, R.G.; Quan, H.N.
 J. Bacteriol. 176, 307-319, 1994
 A:Title: FptA, the Fe(III)-pyochelin receptor of *Pseudomonas aeruginosa*: a phenolate siderophore
 A:Reference number: A36942; MUID:94117363; PMID:8288523
 A:Accession: A36942
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-720 <ANK>
 A:Cross-references: GB:U03161; NID:9454352; PIDN:AAC43213.1; PID:9454353
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, L.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: G83117
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-720 <STO>
 A:Cross-references: GB:AE004839; GB:AE004091; NID:99950434; PIDN:AAG07609.1; GSPDB:GN00176
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: fptA; PA4221
 C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homology <TNN>
 F:98-227/Domain: tonB-dependent receptor amino-terminal homology <TNN>
 F:446-720/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 1.1%; Score 8; DB 2; Length 720;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 VNLVRKRP 177
 |||||
 Db 176 VNLVRKRP 183

RESULT 11

A50187
 probable iron-siderophore receptor YPO1537 [imported] - *Yersinia pestis* (strain CO92)
 C:Species: *Yersinia pestis*
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C:Accession: A50187
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A50187
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-753 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90360.1; PID:gl5979580; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO1537
 C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homology

Query Match 1.1%; Score 8; DB 2; Length 753;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 NLVRKRP 178
 |||||
 Db 199 NLVRKRP 206

RESULT 12

T30917
 hypothetical protein D1007.15 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T30917
 R;Davidson, S.; Rohlfing, T.
 submitted to the EMBL Data Library, August 1999
 A:Description: The sequence of C. elegans cosmid D1007.
 A:Reference number: Z20934
 A:Accession: T30917
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-775 <DAV>
 A:Cross-references: EMBL:AF003151; PIDN:AAB54226.2
 A:Experimental source: strain Bristol N2; clone D1007
 C:Genetics:
 A:Map position: I
 A:Introns: 44/3; 146/3; 224/1; 343/2; 422/3; 494/3; 579/3; 668/3; 709/3
 A:Note: D1007.15
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein D1007.15

Query Match 1.1%; Score 8; DB 2; Length 775;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 IPQSVSII 80
 |||||
 Db 262 IPQSVSII 269

RESULT 13

A40601
 ferrityoverdine receptor - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
 C:Accession: A40601; S54000; F04227
 R;Poole, K.; Neshat, S.; Krebs, K.; Heinrichs, D.E.
 J. Bacteriol. 175, 4597-4604, 1993
 A:Title: Cloning and nucleotide sequence analysis of the ferrityoverdine receptor gene ftyA
 A:Reference number: A40601; MUID:9328663; PMID:8335619
 A:Accession: A40601
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-813 <POO>
 A:Cross-references: GB:L10210; NID:G294360; PIDN:AAA25819.1; PID:G294361
 R;Lamont, I.L.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S53998
 A:Accession: S54000

A:Molecule type: DNA
A:Residues: 1-715,'Y','717-744','YN','745-747','S','749-813 <LAN>
A:Cross-references: EMBL:U07059; NID:G1633044; PIDN:AA60199.1; PID:G466459
R:McMorran, B.J.; Merriman, M.E.; Rombel, I.T.; Lamont, I.L.
Gene 176, 55-59, 1996
A>Title: Characterisation of the pvdE gene which is required for pyoverdine synthesis in
A:Reference number: J05090; MUID:97075909; PMID:8918232
A:Accession: PC4227
A:Molecule type: DNA
A:Residues: 1-49 <MCW>
A:Cross-references: GB:U07359
C:Genetics:
A:Gene: Fpva
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom
C:Keywords: membrane protein
F.184-317/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
F.538-813/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 1.1%; Score 8; DB 2; Length 813;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 GAVYDLND 502
|||||
Db 574 GAVYDLND 581

RESULT 14
H83345
Ferripyoverdine receptor PA2398 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83345
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lapidis, K.; Lm,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H83345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-815 <STO>
A:Cross-references: GB:AE004666; GB:AE004091; NID:G9948438; PIDN:AA605786.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: Fpva, PA2398
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom

Query Match 1.1%; Score 8; DB 2; Length 815;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 GAVYDLND 502
|||||
Db 574 GAVYDLND 581

RESULT 15
E87759
protein D1007.15 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: E87759
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see webites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_el
A:Accession: E87759
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-837 <STO>
A:Cross-references: GB:chr_1; PIDN:AA654232.1; PID:G2088812; GSPDB:GN00019; CESP:DI007.1
C:Genetics:
A:Gene: D1007.15
A:Map position: 1
C:Superfamily: Caenorhabditis elegans hypothetical protein D1007.15

Query Match 1.1%; Score 8; DB 2; Length 837;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 IPQSVSII 80
|||||
Db 324 IPQSVSII 331

RESULT 16
A89959
hypothetical protein SA1562 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: A89959
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: A89959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1274 <KUR>
A:Cross-references: GB:BA000018; PID:G13701536; PIDN:BA642830.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA1562

Query Match 1.1%; Score 8; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 LAPYNGLP 268
|||||
Db 1003 LAPYNGLP 1010

RESULT 17
AD2109
hypothetical protein aar2427 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2109
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2109
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-78 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA674126.1; PID:G17131519; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: aar2427

Query Match 1.0%; Score 7; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 GLSLIGG 466
|||||

Query Match 1.0%; Score 7; DB 2; Length 112;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLA 20
 |||||
 15 AATVLA 21

RESULT 23
 S29304
 hypochetrical protein PA5055 [imported] - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa
 C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 31-Dec-2000
 C:Accession: S29304; B83013; S28376
 R:Timm, A.; Steinhuechel, A.
 Eur. J. Biochem. 209, 15-30, 1992
 A>Title: Cloning and molecular analysis of the poly(3-hydroxyalkanoic acid) gene locus C
 A:Reference number: S29303; MUID:93011120; PMID:1396693
 A:Accession: S29304
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-123 <TIM>

A:Cross-references: EMBL:X66592; NID:g45388; PIDN:CAA47150.1; PID:g45390
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lam,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: B83013
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-123 <STO>

A:Cross-references: GB:AE004919; GB:AE004091; NID:g9951346; PIDN:AAG08440.1; GSPDB:GN001
 A:Experimental source: strain PA01
 A:Genetics:
 A:Gene: PA5055
 C:Superfamily: Pseudomonas aeruginosa hypothetical protein 2 (phact 5' region)

Query Match 1.0%; Score 7; DB 2; Length 123;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 482 LHKASKT 488
 |||||
 DB 9 LHKASKT 15

RESULT 24
 T38756

60s ribosomal protein L32 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C:Accession: T38756
 R:Olliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, September 1996

A:Reference number: Z21809

A:Accession: T38756

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-127 <OLI>

A:Cross-references: EMBL:Z29296; PIDN:CAB16594.1; GSPDB:GN00066; SPDB:SPAC3H5.10

A:Experimental source: strain 972h-; cosmid c3H5

C:Genetics:

A:Gene: SPDB:SPAC3H5.10

A:Map position: 1

C:Superfamily: rat ribosomal protein L32

Query Match 1.0%; Score 7; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 NNKTRY 575
 |||||
 DB 56 NNKTRY 62

RESULT 25
 T39562

60s ribosomal protein L32 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C:Accession: T39562

R:Purnelle, B.; Goffeau, A.; Wood, V.; Lyne, M.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 1998

A:Reference number: Z21863

A:Accession: T39562

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-127 <PUP>

A:Cross-references: EMBL:AL021767; PIDN:CAA16918.1; GSPDB:GN00067; SPDB:SPBC16C6.11

A:Experimental source: strain 972h-; cosmid c16C6

C:Genetics:

A:Gene: SPDB:SPBC16C6.11

A:Map position: 2

C:Superfamily: rat ribosomal protein L32

Query Match 1.0%; Score 7; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 NNKTRY 575
 |||||
 DB 56 NNKTRY 62

RESULT 26
 A84120

Arp synthase protein BH3761 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: A84120

R:Takami, H.; Nakasone, K.; Takaki, Y.; Meno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000

A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: A84120

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-130 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAE07480.1; GSPDB:GN001

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3761

Query Match 1.0%; Score 7; DB 2; Length 130;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 632 NLTMYQ 638
 |||||
 DB 47 NLTMYQ 53

RESULT 27
 S71256

ribosomal protein L27a, cytosolic - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000

C:Accession: S71256

R:Greller, F.; Cooke, R.; Landie, M.; Raynal, M.; Delseny, M.

submitted to the EMBL Data Library, October 1995

A:Reference number: S71253

A:Accession: S71256

A:Molecule type: mRNA

A;Residues: 1-146 <GRE>
A;Cross-references: EMBL:X91959; NID:g1107486; PIDN:CAA63025.1; PID:g1107487
C;Superfamily: rat ribosomal protein L27a
C;Keywords: protein biosynthesis; ribosome

Query Match 1.0%; Score 7; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKVGMRY 318
|||||
Db 54 GKVGMRY 60

RESULT 28
C66732
transcription regulator [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 01-Mar-2002
C;Accession: C66732
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A;Reference number: AB6625; MUID:21235186; PMID:11337471
A;Accession: C66732
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-148 <STO>
A;Cross-references: GB:AE005176; PID:g12723787; PIDN:AAK04957.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ahrC
C;Superfamily: Escherichia coli arginine repressor argR

Query Match 1.0%; Score 7; DB 2; Length 148;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 364 KALAFDA 370
|||||
Db 77 KALAFDA 83

RESULT 29
R6BY29
ribosomal protein L27a.e, cytosolic - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein G3101; protein YGL103w; ribosomal protein YL29
C;Species: Saccharomyces cerevisiae
C;Date: 25-Feb-1985 #sequence_revision 30-Jun-1992 #text_change 21-Jul-2000
C;Accession: A02782; S45696; S64110
R;Kaeuffer, N.F.; Fried, H.M.; Schwindinger, W.F.; Jasin, M.; Warner, J.R.
Nucleic Acids Res. 11, 3123-3135, 1983
A;Title: Cycloheximide resistance in yeast: the gene and its protein.
A;Reference number: A02782; MUID:83220732; PMID:8304624
A;Accession: A02782
A;Molecule type: DNA
A;Residues: 1-149 <KAU>
A;Cross-references: EMBL:X01573; NID:g3619; PIDN:CAA25729.1; PID:g3620
A;Note: a mutation resulting in the replacement of Gln by Glu at position 38 confers res
R;Schwindinger, W.F.; Warner, J.R.
J. Biol. Chem. 262, 5690-5695, 1987
A;Title: Transcriptional elements of the yeast ribosomal protein gene CYH2.
A;Reference number: S45696; MUID:87194763; PMID:3553182
A;Accession: S45696
A;Molecule type: DNA
A;Residues: 1-16 <SCH>
A;Cross-references: EMBL:M19490; NID:g172481; PIDN:AAA35002.1; PID:g553140
R;Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64071
A;Accession: S64110
A;Molecule type: DNA
A;Residues: 1-149 <RIE>

A;Cross-references: EMBL:Z72625; NID:g1322642; PIDN:CAA96808.1; PID:g1322643; GSPDB:GN00
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:CYH2; RPL29; MIPS:YGL103W
A;Cross-references: SGD:S0003071; MIPS:YGL103W
A;Map position: 7L
A;introns: 17/1
C;Superfamily: rat ribosomal protein L27a
C;Keywords: protein biosynthesis; ribosome
F;2-149/Product: ribosomal protein L27a.e #status experimental <MAT>

Query Match 1.0%; Score 7; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKVGMRY 318
|||||
Db 54 GKVGMRY 60

RESULT 30
S77384
hypothetical protein sll1455 - Synecchocystis sp. (strain PCC 6803)
C;Species: Synecchocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S77384
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, M.;
O. K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S77384
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-155 <KAN>
A;Cross-references: EMBL:D90906; GB:AB001339; NID:g1652492; PIDN:BAAL17487.1; PID:g165256
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: Synecchocystis hypothetical protein sll1455

Query Match 1.0%; Score 7; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 NKKLPSL 278
|||||
Db 103 NKKLPSL 109

RESULT 31
AH3359
probable competence-damage protein [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C;Accession: AH3359
R;DeIvecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AH3359
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-165 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL52043.1; PID:g17982810; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0862
A;Map position: I
C;Superfamily: Aquifex aeolicus conserved hypothetical protein aq_1996

Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ATVTAAAL 21
DB 149 ATVTAAAL 155

RESULT 32

T10313
hypohectical protein 44 - *Orygia pseudotsugata* nuclear polyhedrosis virus
C:Species: *Orygia pseudotsugata* nuclear polyhedrosis virus, OpNPV
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10313
R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohmann, G.F.
Virology 229, 381-399, 1997
A:Title: The sequence of the *Orygia pseudotsugata* multicomponent nuclear polyhedrosis
A:Reference number: Z17011; MUID:97271300; PMID:9126251
A:Accession: T10313
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-166 <AHR>
A:Cross-references: EMBL:U75930; NID:g2934903; PIDN:AAC59043.1; PID:g1911290

Query Match 1.0%; Score 7; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLA 20
DB 105 AATVLA 111

RESULT 33

P83062
hypohectical protein PA4661 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83062
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lathig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83062
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-173 <STO>
A:Cross-references: GB:AE004880; GB:AE004091; NID:g9950912; PIDN:AAG08048.1; GSPDB:GN001
A:Experimental source: strain PA01
A:Genetics:
A:Gene: PA4661

Query Match 1.0%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 VLAALSS 23
DB 9 VLAALSS 15

RESULT 34

C86231
hypohectical protein [imported] - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86231
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizier, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shum, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86231
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-174 <STO>
A:Cross-references: GB:AE005172; NID:g2160182; PIDN:AAB60745.1; GSPDB:GN00141
A:Genetics:
A:Map position: 1

Query Match 1.0%; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 DVGSGLN 206
DB 2 DVGSGLN 8

RESULT 35

PMECD
H+-transporting two-sector ATPase (EC 3.6.3.14) delta chain - *Escherichia coli* (strain K
N) Alternate names: hydrogen ion-transporting ATPase delta chain
C:Species: *Escherichia coli*
C:Date: 18-Dec-1981 #sequence_revision 02-Apr-1992 #text_change 01-Mar-2002
C:Accession: A93732; A90101; I41274; H65176; T45005; A01031
R:Gay, N.J.; Walker, J.E.
Nucleic Acids Res. 9, 3919-3926, 1981
A:Title: The *atp* operon: nucleotide sequence of the promoter and the genes for the membr
A:Reference number: A93732; MUID:82059437; PMID:6272190
A:Accession: A93732
A:Molecule type: DNA
A:Residues: 1-177 <GAY>
A:Cross-references: GB:U00264; GB:X00771; NID:g41023; PIDN:CAA23517.1; PID:g41028
R:Mabuchi, K.; Kanazawa, H.; Kayano, T.; Futai, M.
Biochem. Biophys. Res. Commun. 102, 172-179, 1981
A:Title: Nucleotide sequence of the gene coding for the delta subunit of proton-transloc
A:Reference number: A90101; MUID:82068433; PMID:6458296
A:Accession: A90101
A:Molecule type: DNA
A:Residues: 1-81, 'D', 83-177 <MAB>
A:Cross-references: GB:M12213; GB:M12213; NID:g145336; PIDN:AAA20044.1; PID:g145336
R:Kanazawa, H.; Futai, M.
Ann. N. Y. Acad. Sci. 402, 45-64, 1982
A:Title: Structure and function of H+-ATPase: What we have learned from *Escherichia coli*
A:Reference number: I41274; MUID:83176724; PMID:6301339
A:Accession: I41274
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-177 <RES>
A:Cross-references: GB:M25464; NID:g146318; PIDN:AAA83872.1; PID:g146322
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H65176
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-177 <BLAT>
A:Cross-references: GB:AE000450; GB:U00096; NID:g1790166; PIDN:AAC76758.1; PID:g1790173;
A:Experimental source: strain K-12, substrain MG1655
R:Nielsen, J.; Hansen, F.G.; Hoppe, J.; Friedl, P.; Von Meyenburg, K.
Mol. Gen. Genet. 184, 33-39, 1981
A:Title: The nucleotide sequence of the *atp* gene coding for the F-0 subunits a, b, c an
A:Reference number: Z22893; MUID:82147764; PMID:6278247
A:Accession: T45005
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-177 <NIE>
A:Cross-references: EMBL:V00266; NID:g41031; PIDN:CAA23524.1; PID:g41034
C:Genetics:
A:Gene: atpH; uncH
A:Map position: 84 min
C:Complex: this is one of the five chains of the enzymatic component (F1) of the ATPase
C:Superfamily: H⁺-transporting ATP synthase delta chain
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match 1.0%; Score 7; DB 1; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 36
AC0954
A:Title: ATP synthase delta chain [imported] - Salmonella enterica subsp. enterica serovar Typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 17-May-2002
C:Accession: AC0954
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi
A:Reference number: AB0502; PMID:11677608
A:Accession: AC0954
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03127.1; PID:g16504763; GSPDB:GN00176
C:Gene: STY3910
C:Superfamily: H⁺-transporting ATP synthase delta chain

Query Match 1.0%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 37
F86059
A:Title: H⁺-transporting two-sector ATPase delta chain [similarity] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: F86059
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; PMID:21074935; PMID:11206551
A:Accession: F86059
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <STO>
A:Cross-references: GB:AB005174; NID:gl2518592; PIDN:AAG58938.1; GSPDB:GN00145; UWGP:Z52
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: atpH
C:Superfamily: H⁺-transporting ATP synthase delta chain

Query Match 1.0%; Score 7; DB 2; Length 177;

Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 38
AH0500
A:Title: H⁺-transporting two-sector ATPase (EC 3.6.3.14) delta chain [imported] - Yersinia pestis
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
C:Accession: AH0500
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, N.; Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AH0500
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC93573.1; PID:gl5982013; GSPDB:GN00175
C:Genetics:
A:Gene: atpH
C:Superfamily: H⁺-transporting ATP synthase delta chain
C:Keywords: hydrolase

Query Match 1.0%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 39
E91213
A:Title: H⁺-transporting two-sector ATPase (EC 3.6.3.14) delta chain [similarity] - Escherichia coli
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Jun-2002
C:Accession: E91213
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: E91213
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA038100.1; PID:gl3364152; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs4677
C:Superfamily: H⁺-transporting ATP synthase delta chain
C:Keywords: hydrolase

Query Match 1.0%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 40
D82490
A:Title: Hypothetical protein VCA0195 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C.Species: *Vibrio cholerae*
C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #ext_change 02-Feb-2001
C.Accession: D82490
R.Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, R.; Vamathevan, J.; Baas, S.; Qin, H.; Dragoi, I.; Sellers, F.
L. R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A.Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A.Reference number: A82035; MUID:20406833; PMID:10952301
A.Accession: D82490
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-182 <HE1>
A.Cross-references: GB:AE004359; GB:AE003853; NID:g9657575; PIDN:AAF96108.1; GSPDB:GN001
A.Experimental source: serogroup O1, strain N16961; biotype El Tor
C.Genetics:
A.Gene: VCA0195
A.Map position: 2

Query Match 1.0%; Score 7; DB 2; Length 182;
Best Local Similarity 100.0%; Pred.No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LYARGGL 411
|||||
Db 107 LYARGGL 113

RESULT 41
T04031
hypothetical protein F17A8.180 - *Arabidopsis thaliana*
C.Species: *Arabidopsis thaliana* (mouse-ear cress)
C.Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #ext_change 20-Sep-1999
C.Accession: T04031
R.Beyan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Banroff, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A.Reference number: Z15184
A.Molecule type: DNA
A.Accession: T04031
A.Residues: 1-191 <BEV>
A.Cross-references: EMBL:AL049482
A.Experimental source: cultivar Columbia; BAC clone F17A8
C.Genetics:
A.Map position: 4
A.Introns: 53/3; 80/3
A.Note: F17A8.180
C.Superfamily: *Arabidopsis* hypothetical protein F17A8.180

Query Match 1.0%; Score 7; DB 2; Length 191;
Best Local Similarity 100.0%; Pred.No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 SLSQLYT 515
|||||
Db 29 SLSQLYT 35

RESULT 42
E95072
nitroreductase family protein [imported] - *Streptococcus pneumoniae* (strain TIGR4)
C.Species: *Streptococcus pneumoniae*
C.Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 03-Aug-2001
C.Accession: E95072
R.Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Kadane, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A.Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
A.Reference number: A95000; MUID:2157209; PMID:11463916
A.Accession: E95072
A.Status: preliminary
A.Molecule type: DNA

A.Residues: 1-201 <KUR>
A.Cross-references: GB:AE005672; PIDN:AAK74774.1; PID:g14972096; GSPDB:GN00164; TIGR:SP4
A.Experimental source: strain TIGR4
C.Genetics:
A.Gene: SP0622

Query Match 1.0%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred.No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 ARVGGAN 703
|||||
Db 92 ARVGGAN 98

RESULT 43
B97940
nitroreductase (EC 1.-.-.-) [imported] - *Streptococcus pneumoniae* (strain R6)
C.Species: *Streptococcus pneumoniae*
C.Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #ext_change 02-Nov-2001
C.Accession: B97940
R.Hoshing, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mchren, S.; P
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A.Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaaskunas, S.R.;
A.Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
A.Reference number: A97872; MUID:21429245; PMID:11544234
A.Accession: B97940
A.Status: preliminary
A.Molecule type: DNA
A.Accession: B97940
A.Residues: 1-201 <KUR>
A.Cross-references: GB:AE007317; PIDN:AAK99350.1; PID:g15458122; GSPDB:GN00174
C.Genetics:
A.Gene: nrd
C.Keywords: oxidoreductase

Query Match 1.0%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred.No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 ARVGGAN 703
|||||
Db 92 ARVGGAN 98

RESULT 44
B86874
hypothetical protein yuhE [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)
C.Species: *Lactococcus lactis* subsp. *lactis*
C.Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #ext_change 03-Aug-2001
C.Accession: B86874
R.Bolotin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A.Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* s
A.Reference number: A86625; MUID:21235186; PMID:11337471
A.Accession: B86874
A.Status: preliminary
A.Molecule type: DNA
A.Accession: B86874
A.Residues: 1-208 <STO>
A.Cross-references: GB:AE005176; PID:g12725039; PIDN:AAK06092.1; GSPDB:GN00146
A.Experimental source: strain IL1403
C.Genetics:
A.Gene: yuhE

Query Match 1.0%; Score 7; DB 2; Length 208;
Best Local Similarity 100.0%; Pred.No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 645 IGGGVNA 651
|||||
Db 179 IGGGVNA 185

RESULT 45
AG2841
outer membrane protein omp [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AG2841
R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG2841
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43149.1; PID:gl7740625; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: omp
A:Map position: circular chromosome

Query Match 1.0%; Score 7; DB 2; Length 209;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 GSVGRGV 215
DB 117 GSVGRGV 123

Search completed: December 18, 2002, 06:59:29
Job time : 40.351 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:53:55 ; Search time 10.2194 Seconds

(without alignments)
2930.303 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 722
Sequence: 1 MGQFMSVFRIMNTATVLA.....NTFNIGSERTWANTLRYSF 722

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	629	1 DXS_BACHD	Q9K971 bacillus ha
2	8	1.1	197	1 RNM2_PASMU	P57986 pasteurella
3	8	1.1	532	1 ICEP_HSVB	P28990 equine herp
4	8	1.1	644	1 LEU1_MYCTU	P96420 mycobacteri
5	8	1.1	702	1 FOXA_SALTY	O56145 salmonella
6	8	1.1	720	1 FPTA_PSEAE	P42512 pseudomonas
7	8	1.1	815	1 FPVA_PSEAE	P48632 pseudomonas
8	7	1.0	102	1 THIO_CHLMU	P79015 schizosacch
9	7	1.0	127	1 R32A_SCHPO	O42335 schizosacch
10	7	1.0	139	1 VG08_BPMU	Q38480 bacteriopho
11	7	1.0	146	1 R27B_ARATH	O91433 arabidopsis
12	7	1.0	146	1 R27C_ARATH	P49637 arabidopsis
13	7	1.0	148	1 RL28_YEAST	P02406 saccharomyc
14	7	1.0	149	1 RL2A_ERYGR	P78987 erysihe gr
15	7	1.0	166	1 Y024_NPVOP	O10299 ergyia pseu
16	7	1.0	177	1 ATPD_ECOLI	P00831 escherichia
17	7	1.0	210	1 RGSB_CHICK	O9P940 gallus gall
18	7	1.0	210	1 RGSB_MOUSE	O9G420 mus musculu
19	7	1.0	211	1 ROEB_RHILV	Q52866 rhizobium 1
20	7	1.0	214	1 ETS_MOUSE	P48299 mus musculu
21	7	1.0	217	1 OMPW_VIBCH	P17266 vibrio chol
22	7	1.0	256	1 SPSA_BACSU	P39621 bacillus su
23	7	1.0	268	1 ISPE_AQUAE	O67060 aquifex aeo
24	7	1.0	286	1 TEHB_HAEIN	P45134 haemophilus
25	7	1.0	295	1 NADC_STNY3	P74501 haemophilus
26	7	1.0	318	1 FEPB_ECOLI	P14609 escherichia
27	7	1.0	333	1 TRPD_PASMU	P57856 pasteurella
28	7	1.0	334	1 OTCC_HAEIN	P44770 haemophilus
29	7	1.0	344	1 YAAA_SCHPO	O09801 schizosacch
30	7	1.0	351	1 PHOE_CITFR	O01605 citrobacter
31	7	1.0	354	1 MURB_TREPA	O83128 treponema p
32	7	1.0	363	1 V363_PSPB7	P23164 african swi
33	7	1.0			

34	7	1.0	382	1 IPAC_SHIDY	Q03946 shigella dy
35	7	1.0	382	1 IPAC_SHIFL	P18012 shigella fl
36	7	1.0	395	1 BNS1_HUMAN	P05423 homo sapien
37	7	1.0	411	1 VGLG_HSVB	P28967 equine herp
38	7	1.0	411	1 VGLX_HSVB	P32514 equine herp
39	7	1.0	412	1 GLUP_BRUBA	O44623 brucella ab
40	7	1.0	427	1 Y118_TREPA	O83155 treponema p
41	7	1.0	431	1 POLC_HAEIN	P43175 haemophilus
42	7	1.0	443	1 UVSH_EMENT	O02398 emericella
43	7	1.0	454	1 SR54_AQUAE	O67615 aquifex aeo
44	7	1.0	456	1 EPIA_DICDI	P18624 dictyosteli
45	7	1.0	457	1 TBBI_PORPU	P50259 porphyra pu
46	7	1.0	477	1 XYNA_STRLI	P26514 streptomyce
47	7	1.0	485	1 CAT1_NICPL	P49315 nicotiana p
48	7	1.0	506	1 TCNO_PETCR	O43033 petroselinu
49	7	1.0	506	1 Y213_MYCPV	P75478 mycoplasma
50	7	1.0	510	1 C304_DROME	O99917 drosophila
51	7	1.0	522	1 ACCD_SPTOL	O9m317 spinacia ol
52	7	1.0	554	1 HAP4_YEAST	P14064 saccharomyc
53	7	1.0	557	1 C791_SORBI	O43135 sorghum bic
54	7	1.0	574	1 CO9_ONCMY	P06682 oncorhynch
55	7	1.0	586	1 CO9_FUGRU	P79755 fugu rubrip
56	7	1.0	632	1 YKG5_CABEL	P46555 caenorhabdi
57	7	1.0	639	1 PS02_YEAST	P30620 saccharomyc
58	7	1.0	661	1 RED1_RAT	P51400 rattus norv
59	7	1.0	714	1 Y1L2_YEAST	P47025 saccharomyc
60	7	1.0	716	1 HEP4_HSVB	P28946 equine herp
61	7	1.0	741	1 RED1_HUMAN	P78563 homo sapien
62	7	1.0	747	1 GUND_CELFI	P50400 cellulomona
63	7	1.0	759	1 Y363_METUA	O57809 methanococ
64	7	1.0	778	1 RHP9_SCHPO	P87074 schizosacch
65	7	1.0	809	1 PUBB_PSEPU	P87047 pseudomonas
66	7	1.0	811	1 YB03_SCHPO	P87145 schizosacch
67	7	1.0	819	1 PUPA_PSEPU	P25184 pseudomonas
68	7	1.0	823	1 SP97_YEAST	P38863 saccharomyc
69	7	1.0	867	1 EP2_BLAHO	O17152 blastocysti
70	7	1.0	960	1 DLG1_DROME	P31007 drosophila
71	7	1.0	963	1 UBPA_HUMAN	O13107 homo sapien
72	7	1.0	985	1 REVI_YEAST	P12689 saccharomyc
73	7	1.0	1239	1 NME3_MOUSE	O01098 mus musculu
74	7	1.0	1284	1 NRE4_DROME	O94887 drosophila
75	7	1.0	1388	1 HRP3_SCHPO	O14139 schizosacch
76	7	1.0	1514	1 YD34_YEAST	O05471 saccharomyc
77	7	1.0	1643	1 RRP0_NMV	P15095 narcissus m
78	7	1.0	1696	1 ITN2_HUMAN	O9nm33 homo sapien
79	7	1.0	2014	1 YU07_YEAST	P39226 saccharomyc
80	7	1.0	2164	1 POLG_HRV89	P07210 human thino
81	6	0.8	44	1 RL34_RICCN	O92836 rickettsia
82	6	0.8	63	1 YLM4_CABEL	P34407 caenorhabdi
83	6	0.8	65	1 YIME_BPHPI	P10437 bacteriopho
84	6	0.8	70	1 RL31_ECOLI	P02432 escherichia
85	6	0.8	72	1 PSBH_ARATH	P56780 arabidopsis
86	6	0.8	76	1 STBC_ECOLI	P11905 escherichia
87	6	0.8	82	1 YBDJ_ECOLI	O11775 caenorhabdi
88	6	0.8	87	1 YOK2_CABEL	P44708 haemophilus
89	6	0.8	90	1 YER9_HAEIN	P49779 haemophilus
90	6	0.8	93	1 YQHV_BACSU	P04670 autotriapha
91	6	0.8	94	1 VP10_NPVAC	P04691 bacteriopho
92	6	0.8	100	1 REGN_BPP22	P19330 beet necrot
93	6	0.8	101	1 Y12K_BNYVG	O84544 chlamydia t
94	6	0.8	102	1 THIO_CHLTR	P18836 oxytricha t
95	6	0.8	103	1 H4_OXYNO	P47841 oxytrichus
96	6	0.8	103	1 MGE_RABIT	O9y6h6 homo sapien
97	6	0.8	103	1 MIR2_HUMAN	O9w7w2 mus musculu
98	6	0.8	103	1 MIR2_MOUSE	P80739 euplotes cr
99	6	0.8	106	1 H4_EUPCR	O9j1v7 rattus norv
100	6	0.8	107	1 MIR2_RAT	O31494 bacillus su
101	6	0.8	109	1 YDGF_BACSU	P52107 escherichia
102	6	0.8	110	1 CSCC_ECOLI	O2618 methanobac
103	6	0.8	111	1 Y518_METTH	O928w1 bacillus ha
104	6	0.8	113	1 Y655_BACHD	P57915 pasteurella
105	6	0.8			
106	6	0.8	119	1 RNPA_PASMU	

107	120	6	0.8	120	1	R22A_YEAST	P05749	saccharomyc	180	6	0.8	169	1	CRG2_RANTE	P02531	rana tempor
108	121	6	0.8	121	1	R22B_YEAST	P56628	saccharomyc	181	6	0.8	175	1	VS10_ROTBN	P08434	bovine rota
109	122	6	0.8	122	1	RS6E_ARCFU	O29739	archaeoglob	182	6	0.8	175	1	VS10_ROTBN	P04513	bovine rota
110	123	6	0.8	123	1	LY6E_MOUSE	Q64253	mus musculus	183	6	0.8	175	1	VS10_ROTBN	P30032	human rotav
111	124	6	0.8	124	1	RR11_CVAPA	P48136	cyanophora	184	6	0.8	175	1	VS10_ROTBN	P01535	human rotav
112	125	6	0.8	125	1	RR11_WARPO	P06364	marcantiia	185	6	0.8	175	1	VS10_ROTBN	Q06235	mycobacteri
113	126	6	0.8	126	1	RR11_MESVI	Q9muu9	mesocistima	186	6	0.8	177	1	NR13_COTJA	Q90343	coturnix co
114	127	6	0.8	127	1	RR11_ODOSI	P41630	odontella s	187	6	0.8	181	1	BCCP_ANASP	Q06881	anabaena ol
115	128	6	0.8	128	1	RR11_PINTH	P51294	porphyra pu	188	6	0.8	181	1	THIM_SPIOL	P07591	spinacia ol
116	129	6	0.8	129	1	RR11_PORPU	O24709	synchococc	189	6	0.8	182	1	MCRC_METVO	P11567	methanococc
117	130	6	0.8	130	1	RS11_SYNPE	P31490	yersinia en	190	6	0.8	183	1	RL18_CICAR	O65729	cicer ariet
118	131	6	0.8	131	1	YERA_YEREN	P31491	yersinia pe	191	6	0.8	183	1	RRP_DEIRA	Q5rue2	deinococcus
119	132	6	0.8	132	1	YERA_YERPE	Q46864	escherichia	192	6	0.8	183	1	TR13_HUMAN	Q15645	homo sapien
120	133	6	0.8	133	1	YGIT_ECOLI	P06506	spinacia ol	193	6	0.8	183	1	YBCL_ECOLI	P77368	escherichia
121	134	6	0.8	134	1	ATP2_BACFI	P22475	bacillus fi	194	6	0.8	186	1	VMA2_TRTV	P33494	turkey thin
122	135	6	0.8	135	1	EXD2_HELPY	O25987	helicobacte	195	6	0.8	187	1	RL18_ARATH	P42791	arabidopsis
123	136	6	0.8	136	1	YH35_SYNPE	Q31752	synchococc	196	6	0.8	187	1	RL18_HUMAN	Q07020	homo sapien
124	137	6	0.8	137	1	RS12_MYCPU	Q98qde	mycoplasma	197	6	0.8	187	1	RL18_ICTPU	Q90yvo	ictalurus p
125	138	6	0.8	138	1	RR11_ARATH	P56802	arabidopsis	198	6	0.8	187	1	RL18_MOUSE	P35980	mus musculus
126	139	6	0.8	139	1	RR11_PEA	P06587	pisum sativ	199	6	0.8	187	1	RL18_OREMO	Q91836	oreochromis
127	140	6	0.8	140	1	RR11_SPIOL	P06506	spinacia ol	200	6	0.8	187	1	RL18_RAT	P12001	rattus norv
128	141	6	0.8	141	1	RR11_TOBAC	P06365	nicotiana t	201	6	0.8	187	1	RL18_XENLA	P02412	xenopus lae
129	142	6	0.8	142	1	VGAM_LAMB2	P03702	bacterioph	202	6	0.8	187	1	RLNA_XENLA	P09897	xenopus lae
130	143	6	0.8	143	1	RBS_CVLN	P24683	cylindrothe	203	6	0.8	187	1	SURE_TREPA	O83434	treponema p
131	144	6	0.8	144	1	RBS_DETGO	O98946	detonula co	204	6	0.8	188	1	SLP_ECOLI	P37194	escherichia
132	145	6	0.8	145	1	RBS_ODOSI	P49521	odontella s	205	6	0.8	189	1	INAS_MOUSE	P07349	mus musculus
133	146	6	0.8	146	1	RBS_THANO	O98948	thalassiosi	206	6	0.8	190	1	INA2_MOUSE	P01573	mus musculus
134	147	6	0.8	147	1	CNAR_HUMAN	Q04762	homo sapien	207	6	0.8	190	1	INA9_MOUSE	P09235	mus musculus
135	148	6	0.8	148	1	Y085_ARCFU	Q30151	archaeoglob	208	6	0.8	192	1	Y56A_THEME	P58008	thermotoga
136	149	6	0.8	149	1	YSOD_SULAC	P37820	sulfolobus	209	6	0.8	193	1	RL18_TRYBB	P50885	trypanosoma
137	150	6	0.8	150	1	FER_THEAC	P00218	thermoplasm	210	6	0.8	195	1	PYRE_SULTO	Q970X1	sulfolobus
138	151	6	0.8	151	1	HEAC_HOPLI	R82315	hoplosternu	211	6	0.8	197	1	RUVA_BORBU	P70827	borrelia bu
139	152	6	0.8	152	1	YN77_RHIME	Q9X714	rhizobium m	212	6	0.8	198	1	SDH3_YEAST	P33421	saccharomyc
140	153	6	0.8	153	1	IL3_BOVIN	P49875	bos taurus	213	6	0.8	199	1	SYS_CRIGR	P26636	cricketulus
141	154	6	0.8	154	1	RL11_STEIN	P36258	streptomyce	214	6	0.8	200	1	RBFA_ARATH	O65693	arabidopsis
142	155	6	0.8	155	1	SODM_BRAFL	P28761	branchiosto	215	6	0.8	203	1	CTFI_MOUSE	Q60753	mus musculus
143	156	6	0.8	156	1	OM25_HUMAN	P57105	homo sapien	216	6	0.8	203	1	CTFI_RAT	Q63086	rattus norv
144	157	6	0.8	157	1	RL2A_OSCOR	O01358	oscheilus br	217	6	0.8	203	1	RR22_HUMAN	Q92737	homo sapien
145	158	6	0.8	158	1	IL3_SHEEP	Q06435	ovis aries	218	6	0.8	204	1	YQF7_CABEL	Q09271	caenorhabdi
146	159	6	0.8	159	1	RL2A_MOUSE	P14115	mus musculus	219	6	0.8	206	1	OM25_RAT	O9WV14	rattus norv
147	160	6	0.8	160	1	RL2A_RAT	P18445	rattus norv	220	6	0.8	207	1	UCRI_CHRVI	O31214	chromatium
148	161	6	0.8	161	1	RL2A_XENLA	P47830	xenopus lae	221	6	0.8	208	1	DSBA_SHIFL	P52235	shigella fl
149	162	6	0.8	162	1	CB20_XENLA	P52299	xenopus lae	222	6	0.8	208	1	UPP_ECOLI	P25532	escherichia
150	163	6	0.8	163	1	R28A_SCHPO	P36585	schizosacch	223	6	0.8	208	1	UPP_PASMU	Q9CPL8	pasteurella
151	164	6	0.8	164	1	R28B_SCHPO	P57728	schizosacch	224	6	0.8	209	1	IM23_MOUSE	Q9WV08	mus musculus
152	165	6	0.8	165	1	RL28_NEUCR	P08978	neurospora	225	6	0.8	209	1	IM23_RAT	O35093	rattus norv
153	166	6	0.8	166	1	RL2A_DROME	P41092	drosophila	226	6	0.8	209	1	PYRE_COXBU	Q45918	coxiella bu
154	167	6	0.8	167	1	Y881_VIBCH	Q9Ktm1	vibriochol	227	6	0.8	210	1	RHOB_HUMAN	O00212	homo sapien
155	168	6	0.8	168	1	NAAA_BPT3	P20331	bacterioph	228	6	0.8	211	1	NUSB_ANASP	Q8YWS6	anabaena sp
156	169	6	0.8	169	1	NAAA_BPT7	P00806	bacterioph	229	6	0.8	211	1	PIMT_PSEAE	P45683	pseudomonas
157	170	6	0.8	170	1	YOM1_PHOPR	P29741	photobacter	230	6	0.8	213	1	RB4A_HUMAN	P20338	homo sapien
158	171	6	0.8	171	1	YPR3_ECOLI	P10027	escherichia	231	6	0.8	213	1	RB4A_MOUSE	P56371	mus musculus
159	172	6	0.8	172	1	RS7_BACST	P22744	bacillus st	232	6	0.8	213	1	RB4A_RAT	P05714	rattus norv
160	173	6	0.8	173	1	RS7_BACSU	P21469	bacillus su	233	6	0.8	214	1	YQGE_BACST	P28754	bacillus st
161	174	6	0.8	174	1	RS7_TETH	P17291	thermus the	234	6	0.8	216	1	KAD_PSEPU	Q9Z409	pseudomonas
162	175	6	0.8	175	1	RS7_BACHD	Q9Z918	bacillus ha	235	6	0.8	217	1	HXB7_HUMAN	P09629	homo sapien
163	176	6	0.8	176	1	RS7_LISMO	Q9Z914	listeria mo	236	6	0.8	218	1	CB25_TETTH	P09236	tetrahymena
164	177	6	0.8	177	1	RS7_NEIMA	Q91r15	neisseria m	237	6	0.8	219	1	Y25K_BNVPF	P19229	beet necrot
165	178	6	0.8	178	1	RS7_RALSO	O8XV09	ralstonia s	238	6	0.8	221	1	NUSB_SVNY3	P74395	synchocyst
166	179	6	0.8	179	1	RS7_THICU	O50564	thiobacillu	239	6	0.8	222	1	YMW7_YEAST	Q04272	saccharomyc
167	180	6	0.8	180	1	NIFX_AZOVI	P35642	eikenella c	240	6	0.8	222	1	HXC5_HUMAN	Q00444	homo sapien
168	181	6	0.8	181	1	NIFX_AZOVI	P14887	azotobacte	241	6	0.8	222	1	HXC5_MOUSE	P32043	mus musculus
169	182	6	0.8	182	1	COAD_SVNY3	Q55435	synchocyst	242	6	0.8	223	1	DEOC_PASMU	P57937	pasteurella
170	183	6	0.8	183	1	RS7_RICCN	Q92194	ricchetti	243	6	0.8	223	1	PHOP_ECOLI	P23836	escherichia
171	184	6	0.8	184	1	RS7_RICPR	P41081	rickettsia	244	6	0.8	224	1	PHOP_SALTY	P14146	salmonella
172	185	6	0.8	185	1	YSH2_CAEL	Q09944	caenorhabdi	245	6	0.8	230	1	PIMT_ARATH	Q42539	arabidopsis
173	186	6	0.8	186	1	CHB2_BOMMO	P08828	bombyx mori	246	6	0.8	231	1	CUT2_CABEL	P34682	caenorhabdi
174	187	6	0.8	187	1	YC51_CVAPA	P48327	cyanophora	247	6	0.8	232	1	VSPA_BOTJA	P81661	bothrops ja
175	188	6	0.8	188	1	NOLB_RHIFR	P33208	rhizobium f	248	6	0.8	234	1	PUR7_SULTO	Q970V9	sulfolobus
176	189	6	0.8	189	1	NOLB_RHISN	P55713	rhizobium s	249	6	0.8	236	1	CD63_BOVIN	O9X822	bos taurus
177	190	6	0.8	190	1	COAD_SVNP7	Q55235	synchococc	250	6	0.8	237	1	CD63_HUMAN	P08962	homo sapien
178	191	6	0.8	191	1	ET3_RAT	P13207	rattus norv	251	6	0.8	237	1	CD63_MOUSE	P41731	mus musculus
179	192	6	0.8	192	1	YCF3_GALSU	Q08814	galdieria s	252	6	0.8	237	1	CD63_RABIT	Q28709	oryctolagus

253	6	0.8	237	1	CD63_RAT	P26648	rattus novy	326	6	0.8	273	1	YF00_ECOLI	P53132	escherichia
254	6	0.8	239	1	CLDE_HUMAN	O95500	homo sapien	327	6	0.8	275	1	NHS_STRAP	P53391	streptomyces
255	6	0.8	239	1	CLDE_MOUSE	O95083	mus musculus	328	6	0.8	274	1	A464_CHVP1	O98514	paramesitum
256	6	0.8	239	1	RS2_HAEN	P44371	mus musculus	329	6	0.8	276	1	OCCCT_AGRVU	P35121	agrobacteri
257	6	0.8	240	1	GTO1_MOUSE	O09311	mus musculus	330	6	0.8	276	1	SLBP_DROME	O94966	drosophila
258	6	0.8	241	1	GTO1_HUMAN	P76417	homo sapien	331	6	0.8	277	1	TP2A_NERPE	O94602	aeropyrum p
259	6	0.8	241	1	GTO1_PIG	O94115	sus scrofa	332	6	0.8	278	1	PIRF_CLAPU	O94118	chlamydia m
260	6	0.8	241	1	GTO1_RAT	O94239	rattus novy	333	6	0.8	278	1	Y1A_RHIS	P17986	rhizobium s
261	6	0.8	241	1	YH0H_STRGR	P24222	streptomyces	334	6	0.8	279	1	DAFV_AQUAE	O67693	aquifex aeo
262	6	0.8	242	1	LECI_PSORP	O24113	psophocarpus	335	6	0.8	280	1	KDQA_RHIME	O92499	rhizobium m
263	6	0.8	244	1	ARC3_CBCP	O00901	clostridium	336	6	0.8	280	1	SOL3_YEAST	P38858	saccharomyces
264	6	0.8	245	1	ICP3_HSV1N	P37319	herpes simp	337	6	0.8	280	1	SSM1_YEAST	P53078	saccharomyces
265	6	0.8	245	1	X123_HUMAN	O15884	homo sapien	338	6	0.8	280	1	THRM_ECOS7	P53078	saccharomyces
266	6	0.8	245	1	Y0BG_ECOLI	P32688	escherichia	339	6	0.8	281	1	THRM_ECOLI	P38388	escherichia
267	6	0.8	246	1	RRM2_HUMAN	O94143	homo sapien	340	6	0.8	281	1	Y645_PASNU	O94142	escherichia
268	6	0.8	246	1	SURE_PASNU	P57955	pasteurella	341	6	0.8	282	1	CC08_CAEEL	O94142	escherichia
269	6	0.8	246	1	MECG_SALTY	O82397	salmomella	342	6	0.8	282	1	VARA_PROST	P46117	providencia
270	6	0.8	246	1	MECG_SALTY	P37457	salmomella	343	6	0.8	283	1	YRAL_HAEN	P45298	haemophilus
271	6	0.8	246	1	Y171_AQUAE	O66558	aquifex aeo	344	6	0.8	284	1	STAR_MESAU	P70114	mesocricetus
272	6	0.8	248	1	ARGB_IACPL	O08320	lactobacilli	345	6	0.8	286	1	DMA_HAEN	P44431	haemophilus
273	6	0.8	248	1	H162_CAMJE	O94915	campylobacter	346	6	0.8	286	1	YMB1_CAEEL	O94915	campylobacter
274	6	0.8	248	1	ICP3_HSV1	P36313	herpes simp	347	6	0.8	288	1	ICMT_XENLA	O12947	x protein-s
275	6	0.8	250	1	ARC3_CLOAM	O46134	clostridium	348	6	0.8	290	1	BUDR_KLFTRE	P53666	klebsiella
276	6	0.8	250	1	DLX8_BRARE	O98879	brachydantio	349	6	0.8	290	1	IMP4_YEAST	P53941	saccharomyces
277	6	0.8	250	1	NGN2_HUMAN	O94243	homo sapien	350	6	0.8	290	1	LEC_EAUPU	P16030	baumhina pu
278	6	0.8	250	1	SURE_VIBCH	O94149	vibriol chol	351	6	0.8	290	1	YAS7_METUA	O58457	methanococcus
279	6	0.8	251	1	Y095_HAEN	O57060	haemophilus	352	6	0.8	291	1	THIG_CYACA	O12947	x protein-s
280	6	0.8	251	1	Y137_CHLPN	O92946	chlamydia m	353	6	0.8	292	1	ARGB_DEIRA	O92946	deinococcus
281	6	0.8	252	1	ICP3_HSV1D	P37318	herpes simp	354	6	0.8	294	1	MENA_MYCLE	O07134	mycobacteri
282	6	0.8	252	1	PRGK_SALTY	P41786	salmomella	355	6	0.8	294	1	Y301_METUA	O57149	methanococcus
283	6	0.8	252	1	RS2_PASNU	P57982	pasteurella	356	6	0.8	295	1	MURB_RICCN	O92118	rickettsia
284	6	0.8	253	1	PSA3_SCHPO	O59780	schizosacchar	357	6	0.8	295	1	MURB_RICCN	O92118	rickettsia
285	6	0.8	253	1	SURE_ECOLI	P36664	escherichia	358	6	0.8	296	1	DAFV_MYCLE	P46814	mycobacteri
286	6	0.8	253	1	SURE_SALTY	O84594	salmomella	359	6	0.8	296	1	MURB_CHLMU	O94189	chlamydia m
287	6	0.8	253	1	Y883_MYCTU	O10545	mycobacteri	360	6	0.8	296	1	THTR_BOVIN	P00586	bos taurus
288	6	0.8	254	1	ATTA_TYRIN	P50725	trichoplusi	361	6	0.8	296	1	THTR_CRICR	P46635	cricketus
289	6	0.8	254	1	SURE_YERPE	O82829	yerhina pe	362	6	0.8	296	1	THTR_HUMAN	O16762	homo sapien
290	6	0.8	255	1	SOL4_YEAST	P53115	saccharomyces	363	6	0.8	296	1	YS42_CAEEL	O09311	caenorhabdi
291	6	0.8	255	1	TH14_PYRHO	O59082	pyrococcus	364	6	0.8	297	1	BPHC_BURCE	P47228	buckholderi
292	6	0.8	255	1	VSPA_BOTAT	P04971	bohrrops at	365	6	0.8	297	1	ISPA_BACST	O08291	baecillus st
293	6	0.8	256	1	CB4A_LYCPS	P27524	lycopersico	366	6	0.8	298	1	YSMK_CAEEL	O12948	caenorhabdi
294	6	0.8	256	1	TPIS_AGRYS	O84593	agrobacteri	367	6	0.8	299	1	ARGB_THERN	O87470	thermaer
295	6	0.8	257	1	Y453_AQUAE	O66760	aquifex aeo	368	6	0.8	300	1	NCSR_HUMAN	P00387	homo sapien
296	6	0.8	258	1	RPIA_YEAST	O12189	saccharomyces	369	6	0.8	300	1	P34_RICPR	O92455	rickettsia
297	6	0.8	258	1	SURE_COXBU	O94121	coxijella bu	370	6	0.8	302	1	AROD_PYRAE	O82459	pyrococcus
298	6	0.8	258	1	Y256_MYCPN	P75421	mycoplasma	371	6	0.8	302	1	BPHC_PSEPS	P08695	pseudomonas
299	6	0.8	260	1	ER81_YEAST	P17261	saccharomyces	372	6	0.8	302	1	PPR2_HUMAN	O94915	campylobacter
300	6	0.8	261	1	YK83_CAEEL	P34349	caenorhabdi	373	6	0.8	302	1	PPR2_MOUSE	O70489	rattus novy
301	6	0.8	263	1	ICP3_HSV1F	P08853	herpes simp	374	6	0.8	302	1	PPR2_MOUSE	O70489	rattus novy
302	6	0.8	263	1	PROC_TREPA	P27771	treponema p	375	6	0.8	303	1	URIC_MOUSE	O05587	salmomella
303	6	0.8	263	1	YG3S_YEAST	P53389	saccharomyces	376	6	0.8	303	1	POCR_SALTY	P50177	synecocyst
304	6	0.8	264	1	AMPW_BUCAL	P57324	buchnera ap	377	6	0.8	304	1	Y722_SYNY3	P50177	synecocyst
305	6	0.8	266	1	ETC1_STANU	P01553	staphylococ	378	6	0.8	304	1	T212_LACIC	P28960	equine herp
306	6	0.8	266	1	TH14_METTH	O27657	methanobact	379	6	0.8	304	1	U149_HSVB	O28199	bos taurus
307	6	0.8	267	1	COB2_HUMAN	P27701	homo sapien	380	6	0.8	305	1	ERA_VIBCH	O28199	bos taurus
308	6	0.8	267	1	PM27_STRPU	O26616	strongyloce	381	6	0.8	307	1	CD5R_BOVIN	O15078	homo sapien
309	6	0.8	267	1	PROB_CLOAB	O97663	clostridium	382	6	0.8	307	1	CD5R_HUMAN	O15078	homo sapien
310	6	0.8	267	1	TRBJ_RHISN	P55400	rhizobium s	383	6	0.8	307	1	CD5R_MOUSE	O62938	mus musculus
311	6	0.8	267	1	Y123_THEMA	O94915	thermotoga	384	6	0.8	309	1	VP62_MRDV	P22120	maize rough
312	6	0.8	268	1	CLCR_HUMAN	O94915	homo sapien	385	6	0.8	310	1	UDP_HUMAN	O16831	homo sapien
313	6	0.8	268	1	FA12_RHIME	P58881	rhizobium m	386	6	0.8	311	1	FMT_MYCGE	P47605	mycoplasma
314	6	0.8	268	1	KNH1_YEAST	P50112	saccharomyces	387	6	0.8	311	1	SOX2_XENLA	O42569	xenopus lae
315	6	0.8	268	1	SURE_DEIRA	O94915	deinococcus	388	6	0.8	311	1	YBEX_ECOLI	P41409	escherichia
316	6	0.8	268	1	TRYP_STRGA	O54179	streptomyces	389	6	0.8	312	1	TRUB_BUCAL	P57456	buchnera ap
317	6	0.8	269	1	AQPI_HUMAN	P29972	homo sapien	390	6	0.8	312	1	VG06_BPMLS	O05278	mycobacteri
318	6	0.8	269	1	AQPI_MOUSE	O02013	mus musculus	391	6	0.8	313	1	ARAP_BACBU	P94529	baecillus su
319	6	0.8	269	1	AQPI_RAT	P29975	rattus novy	392	6	0.8	313	1	PENA_ECOLI	O02049	buckholderi
320	6	0.8	270	1	YFCA_ECOLI	P14008	escherichia	393	6	0.8	313	1	YDCU_ECOLI	P77156	escherichia
321	6	0.8	271	1	Y416_THEMA	O94915	thermotoga	394	6	0.8	314	1	YG98_MYCTU	P48430	gallus gall
322	6	0.8	271	1	AQPI_BOVIN	P47665	bos taurus	395	6	0.8	315	1	SOX2_CHICK	P58312	mycobacteri
323	6	0.8	272	1	AQPI_SHEEP	P56401	ovis aries	396	6	0.8	316	1	YS74_CAEEL	O08618	caenorhabdi
324	6	0.8	273	1	YAEZ_YEAST	P29478	saccharomyces	397	6	0.8	317	1	SOX2_HUMAN	P48431	homo sapien
325	6	0.8	273	1	YAEZ_ECOLI	P77206	escherichia	398	6	0.8	317	1	YVCK_BACBU	O06974	baecillus su

399	6	0.8	318	1	TRXB_VIBCH	Q9kss4	vibrio chol	472	6	0.8	357	1	Z212_HUMAN	Q9udv6	homo sapien
400	6	0.8	320	1	ALDR_HORVU	P23901	hordeum vul	473	6	0.8	358	1	A2HS_CAVPO	O70159	cavia porce
401	6	0.8	320	1	MRAM_YERPE	Q8zif7	yersinia pe	474	6	0.8	358	1	CADH_MEDSA	P31656	medicago sa
402	6	0.8	320	1	SOX2_SHEEP	P54231	ovis aries	475	6	0.8	358	1	PHLC_TRYBB	P09194	trypanosoma
403	6	0.8	320	1	Y033_YEAST	Q12117	saccharomyc	476	6	0.8	358	1	RECA_XENBV	P96185	xenorhabdus
404	6	0.8	320	1	FCL_HUMAN	Q13630	homo sapien	477	6	0.8	358	1	RECA_XENNE	Q9x505	xenorhabdus
405	6	0.8	321	1	TRYG_HUMAN	Q9nrr2	homo sapien	478	6	0.8	359	1	MAS2_AGRRH	P27873	agrobacteri
406	6	0.8	323	1	COBD_ARPER	Q9yaa0	aeropyrum p	479	6	0.8	359	1	Y564_MYCLE	Q9ccn9	mycobacteri
407	6	0.8	324	1	YAMI_RHISN	P55568	rhizobium s	480	6	0.8	359	1	Y199_MYCTU	O07733	mycobacteri
408	6	0.8	325	1	CVF_SYPN2	P26293	synthecococc	481	6	0.8	361	1	DCAM_DAUCA	Q9axe3	caenor caro
409	6	0.8	325	1	RECA_PROVU	P26346	proteus vul	482	6	0.8	361	1	GLNA_PANAR	O04831	panulirus a
410	6	0.8	325	1	YBBG_ECOLI	P45394	escherichia	483	6	0.8	362	1	DCAM_IPOBA	Q9m6k1	ipomea bat
411	6	0.8	326	1	UNG2_HUMAN	Q00253	homo sapien	484	6	0.8	362	1	DCAM_PHANI	Q96471	pharbitis n
412	6	0.8	326	1	VS09_ROTBA	P22674	bovine rota	485	6	0.8	363	1	LACK_AGRRD	Q01937	agrobacteri
413	6	0.8	327	1	KDGT_ECOS7	Q8x4q7	escherichia	486	6	0.8	363	1	VP43_NPVAC	P34050	autographa
414	6	0.8	327	1	KDGT_ECOLI	P32172	escherichia	487	6	0.8	363	1	YKL2_CAEEL	P42169	caenorhabdi
415	6	0.8	329	1	PIN2_CAEEL	Q19157	caenorhabdi	488	6	0.8	364	1	E2B1_PYPHO	O58433	pyrococcus
416	6	0.8	330	1	ASNA_HAEIN	P44338	haemophilus	489	6	0.8	369	1	MX_STRPY	P16946	streptococc
417	6	0.8	330	1	ASNA_STRPN	Q97nq0	streptococc	490	6	0.8	370	1	DNRJ_STRPE	P25048	streptomyce
418	6	0.8	330	1	ASNA_STRPY	Q95yu0	streptococc	491	6	0.8	370	1	TRZD_ENTCL	O87589	enterobacte
419	6	0.8	330	1	HBPB_ARATH	P43273	arabidopsis	492	6	0.8	372	1	M22_STRPY	P50469	streptococc
420	6	0.8	331	1	HXAL_MOUSE	P09022	mus musculus	493	6	0.8	373	1	CD14_BOVIN	Q95122	bos taurus
421	6	0.8	331	1	PUR7_ARCFU	O28996	archaeoglob	494	6	0.8	374	1	CARA_XYLFA	Q9pec2	xylella fas
422	6	0.8	331	1	PYRB_PSYTL	Q934t0	psychrobact	495	6	0.8	375	1	ADH4_KLUJA	P49385	kluyveromyc
423	6	0.8	332	1	RUVB_STRPN	Q97sr6	streptococc	496	6	0.8	375	1	FLAB_VIBAN	Q56572	vibrio angu
424	6	0.8	333	1	HXAL_RAT	O08656	rattus norv	497	6	0.8	375	1	TLSX_MYCPN	P75159	mycoplasma
425	6	0.8	333	1	TRPD_HAEIN	P43858	haemophilus	498	6	0.8	375	1	YY10_MYCTU	Q49721	mycobacteri
426	6	0.8	334	1	MDHM_YEAST	P17505	saccharomyc	499	6	0.8	375	1	YV10_MYCTU	Q50716	mycobacteri
427	6	0.8	335	1	GPDA_HAEIN	P43798	haemophilus	500	6	0.8	376	1	HOSC_THETH	O87198	thermus the
428	6	0.8	335	1	HXAL_HUMAN	P49639	homo sapien	501	6	0.8	376	1	SPH_STRPY	P50470	streptococc
429	6	0.8	337	1	GPDA_PASMU	Q9cl17	pasteurella	502	6	0.8	379	1	FTSZ_MYCLE	Q9ccea	mycobacteri
430	6	0.8	339	1	KDGT_BRWCH	P15701	erwinia chr	503	6	0.8	379	1	FTSZ_MYCTU	Q08378	mycobacteri
431	6	0.8	339	1	MOZL_CAEEL	Q9tzm2	caenorhabdi	504	6	0.8	379	1	VAT1_TORCA	P19333	torpedo cal
432	6	0.8	341	1	TRPD_LACCA	P17170	lactobacill	505	6	0.8	381	1	RAPF_BAGSU	P71002	bacillus su
433	6	0.8	342	1	RECA_BRWCA	P26344	erwinia car	506	6	0.8	382	1	YC44_ODOSI	P49539	odontella s
434	6	0.8	342	1	YR22_MYCTU	P26344	erwinia car	507	6	0.8	383	1	TLSX_MYCGE	Q49434	mycoplasma
435	6	0.8	342	1	Y9J2_YEAST	P71691	mycobacteri	508	6	0.8	386	1	FTSZ_MYCKA	Q9kx55	mycobacteri
436	6	0.8	342	1	Y9J2_CAEEL	Q22306	caenorhabdi	509	6	0.8	389	1	YDHP_ECOS7	Q8x625	escherichia
437	6	0.8	343	1	NOV_XENLA	P51609	xenopus lae	510	6	0.8	389	1	YDHP_ECOLI	P77389	escherichia
438	6	0.8	343	1	PDXA_BRUME	P58712	brucella me	511	6	0.8	391	1	NIFV_RHOSH	Q01181	rhodobacter
439	6	0.8	344	1	ALF_MYCTU	O06313	mycobacteri	512	6	0.8	392	1	SERA_MANSE	P14754	manduca sex
440	6	0.8	345	1	HIOM_BOVIN	P10950	bos taurus	513	6	0.8	392	1	YC4P_CAEEL	Q22915	caenorhabdi
441	6	0.8	346	1	GALM_ECOLI	P40681	escherichia	514	6	0.8	392	1	YFGI_ECOLI	P77774	escherichia
442	6	0.8	346	1	RFAF_HAEIN	P45042	haemophilus	515	6	0.8	393	1	DXR_RALSO	Q8x215	raistonia s
443	6	0.8	347	1	ABB2_MOUSE	P98199	mus musculus	516	6	0.8	394	1	BEXD_HAEIN	P22236	haemophilus
444	6	0.8	347	1	GBA5_DICDI	P34043	dictyosteli	517	6	0.8	394	1	MALA_BACST	Q45632	bacillus st
445	6	0.8	347	1	ID12_LACLA	Q9cif5	lactococcus	518	6	0.8	395	1	GPI8_HUMAN	Q92643	homo sapien
446	6	0.8	347	1	Y999_MYCPN	P75593	mycoplasma	519	6	0.8	396	1	DXR_DEIFA	Q9ru84	deinococcus
447	6	0.8	347	1	YREG_HAEIN	P44720	haemophilus	520	6	0.8	396	1	ILJ3_MOUSE	P26952	mus musculu
448	6	0.8	348	1	RECA_VIBAN	P26348	vibrio angu	521	6	0.8	397	1	GALI_TREPA	O83433	treponema p
449	6	0.8	349	1	TL29_ARATH	P82281	arabidopsis	522	6	0.8	398	1	CIW4_MOUSE	O88454	mus musculu
450	6	0.8	350	1	MLF1_MALFU	Q01940	malassezia	523	6	0.8	399	1	FTSZ_STRCO	P45500	streptomyce
451	6	0.8	350	1	XP22_HUMAN	O75695	homo sapien	524	6	0.8	400	1	ACHI_BOMMO	Q01383	bombyx mori
452	6	0.8	351	1	NOV_CHICK	P28686	gallus gall	525	6	0.8	400	1	CINA_ECOLI	P77808	escherichia
453	6	0.8	351	1	YAML_STRLM	P33823	streptomyce	526	6	0.8	401	1	YR07_CAEEL	Q09424	caenorhabdi
454	6	0.8	352	1	RECA_ECOLI	Q93017	escherichia	527	6	0.8	402	1	PGK_HELPJ	Q9zjpl	helicobacte
455	6	0.8	352	1	RECA_RALSO	Q8y1y6	raistonia s	528	6	0.8	402	1	PGK_HELPJ	P56154	helicobacte
456	6	0.8	352	1	RECA_SALTY	Q8xet0	salmonella	529	6	0.8	403	1	S142_RAT	Q99m80	rattus norv
457	6	0.8	352	1	RECA_SHISO	Q9zff6	shigella so	530	6	0.8	405	1	PGK_CORGL	Q01655	corynebacte
458	6	0.8	353	1	LICH_PSEGL	O05490	pseudomonas	531	6	0.8	407	1	AATC_ORYSA	P37833	oryza sativ
459	6	0.8	353	1	NOV_COTJA	P42642	coturnix co	532	6	0.8	407	1	CPXD_AGRTU	P24467	agrobacteri
460	6	0.8	353	1	RECA_AERSA	P94190	aeromonas s	533	6	0.8	407	1	FTSZ_STRGR	P45501	streptomyce
461	6	0.8	353	1	RECA_ALCEU	Q9apb2	alcaligenes	534	6	0.8	408	1	PGK_SULSO	P50317	sulfolobus
462	6	0.8	353	1	RECA_ENTAG	P33037	enterobacte	535	6	0.8	408	1	YBX0_ARATH	Q9zq34	arabidopsis
463	6	0.8	353	1	RECA_SERWA	P17479	serratia ma	536	6	0.8	409	1	IHH_XENLA	O91612	xenopus lae
464	6	0.8	354	1	FENR_CHLRE	P53991	chlamydomon	537	6	0.8	410	1	HXA3_HETFR	Q9ia21	heterodontu
465	6	0.8	354	1	RECA_HAEIN	P43705	haemophilus	538	6	0.8	412	1	ACDS_HUMAN	P16219	homo sapien
466	6	0.8	354	1	RECA_PROMI	P11406	proteus mir	539	6	0.8	412	1	ACDS_MOUSE	Q07417	mus musculu
467	6	0.8	354	1	RECA_VIBCH	P45383	vibrio chol	540	6	0.8	412	1	ACDS_RAT	P15651	rattus norv
468	6	0.8	355	1	EFTS_HELPJ	Q9zj71	helicobacte	541	6	0.8	414	1	YAFU_ECOLI	P04335	escherichia
469	6	0.8	355	1	EFTS_HELPJ	P55975	helicobacte	542	6	0.8	414	1	YGJU_ECOLI	P42602	escherichia
470	6	0.8	355	1	RECA_YERPE	P37858	yersinia pe	543	6	0.8	415	1	RFBX_ECOLI	P37746	escherichia
471	6	0.8	357	1	AROB_STRPY	Q99yr3	streptococc	544	6	0.8	416	1	DAD2_RHILO	Q98b75	rhizobium l

545	6	0.8	417	1	AGP_PROBE	Q52309	providencia	618	6	0.8	469	1	NIFN_RHISN	P56674	rhizobium s
546	6	0.8	417	1	PGKB_CRIFA	P08966	citridhida f	619	6	0.8	470	1	NIFD_METTM	Q50788	methanobact
547	6	0.8	417	1	PGKB_LEIMA	Q27683	leishmania	620	6	0.8	470	1	YICD_ECOLI	P31440	eschierichia
548	6	0.8	417	1	PGKB_LEIME	Q27684	leishmania	621	6	0.8	471	1	GATA_THERM	Q91cxc	thermus the
549	6	0.8	417	1	VGAD_HSVBP	P24906	bovine hepp	622	6	0.8	471	1	VATB_DEIRA	Q9twg7	thermococcus
550	6	0.8	417	1	VGAD_HSVBS	Q08100	bovine hepp	623	6	0.8	472	1	LEU2_BACD	Q9x8f0	baecillus ha
551	6	0.8	419	1	PROA_STRCO	Q9rdk1	streptomyce	624	6	0.8	473	1	BGLB_MICBI	P38645	microdispor
552	6	0.8	422	1	KASI_STRCN	Q02378	streptomyce	625	6	0.8	474	1	BLAR_FELCA	Q9st6c	felis silve
553	6	0.8	424	1	COA_BFPD	P03661	bacteriophia	626	6	0.8	475	1	PEDB_LACIA	P91447	lactococcus
554	6	0.8	424	1	COA_BPM13	P03662	bacteriophia	627	6	0.8	475	1	PPB_SBRMA	P33483	serattia ma
555	6	0.8	424	1	IFZG_ARCFU	Q29663	archaeoglob	628	6	0.8	476	1	PPBH_PSEAE	P33483	pseudomonas
556	6	0.8	424	1	ZEP1_MOUSE	P08042	mus muscullu	629	6	0.8	477	1	HRPB_PSEAE	P31778	raistonia s
557	6	0.8	425	1	GCH2_MYCTU	P71684	mycobacteri	630	6	0.8	477	1	NIFD_METTM	P71526	methanococc
558	6	0.8	426	1	Y958_METUA	Q58468	methanococc	631	6	0.8	477	1	YGUI_ECOLI	P45590	eschierichia
559	6	0.8	427	1	DSL1_CHICK	P34822	gallus galli	632	6	0.8	478	1	GATA_MYCPN	P75534	mycoplasma
560	6	0.8	427	1	IRP3_HUMAN	Q14653	homo sapien	633	6	0.8	478	1	TBG_RTFPI	P54405	reticulomyx
561	6	0.8	428	1	AMY1_ORYSA	P17654	oryza sativ	634	6	0.8	478	1	TIG_AQUAE	Q67358	aquifex aeo
562	6	0.8	428	1	GPRX_ORYLA	Q91178	oryzias lat	635	6	0.8	479	1	PGKC_LEIMA	P50312	leishmania
563	6	0.8	428	1	MTT8_THERM	P29749	thermus the	636	6	0.8	479	1	PGKC_LEIME	Q27685	leishmania
564	6	0.8	428	1	PURA_CLOAB	Q97d87	clostridium	637	6	0.8	479	1	YAH1_HAETN	Q57048	haemophilus
565	6	0.8	430	1	BMP7_MOUSE	P23159	mus muscullu	638	6	0.8	480	1	GLG1_RHIME	P58393	rhizobium m
566	6	0.8	432	1	PROA_DEIRA	Q9rdcd	deinococcus	639	6	0.8	480	1	NIFD_CYAA5	O07642	cynaohece
567	6	0.8	432	1	TOLB_PSEAE	P50601	pseudomonas	640	6	0.8	480	1	YB90_YEAST	P38348	saccharomyc
568	6	0.8	432	1	YKQ7_CAEEL	O16686	caenorhabdi	641	6	0.8	481	1	DXK6_XENIA	P54824	xenopus lae
569	6	0.8	432	1	YKQJ_BACSU	P54444	bacillus su	642	6	0.8	481	1	YAAU_HAETN	P44555	haemophilus
570	6	0.8	434	1	YAA7_XYLFA	Q9pe11	xyellia fas	643	6	0.8	482	1	CG23_SCHPO	P10815	schiizosach
571	6	0.8	437	1	INTR_SACER	P22877	saccharopol	644	6	0.8	482	1	K6B2_HUMAN	Q9ubso	h ribosomal
572	6	0.8	439	1	GFO_ZYMO	Q07982	zymomonas m	645	6	0.8	482	1	SAHH_STRAA	Q93662	streptomyce
573	6	0.8	439	1	PHT3_PSEPU	Q05183	pseudomonas	646	6	0.8	483	1	DXK6_HUMAN	P25196	homo sapien
574	6	0.8	439	1	SLAP_LACHE	P38059	lactobacill	647	6	0.8	483	1	DXK6_MOUSE	P54823	mus muscullu
575	6	0.8	441	1	GAT6_RAT	P46153	rattus norv	648	6	0.8	483	1	PREG_METEX	Q06518	methylobact
576	6	0.8	442	1	FTS2_CORGL	P94337	corynebacte	649	6	0.8	483	1	XYLB_KLEPN	P29444	klebsiella
577	6	0.8	443	1	MCRY_METTH	P21111	methanobact	650	6	0.8	485	1	K6B2_MOUSE	Q941m4	mus muscullu
578	6	0.8	443	1	YAAU_ECOLI	P31679	eschierichia	651	6	0.8	485	1	NAB1_MESAU	Q35589	mesocricetu
579	6	0.8	444	1	GAT6_MOUSE	O61169	mus muscullu	652	6	0.8	485	1	WDRD_HUMAN	Q941n2	homo sapien
580	6	0.8	444	1	GLGC_HAETN	P43796	haemophilus	653	6	0.8	485	1	WDRD_MOUSE	Q91v09	mus muscullu
581	6	0.8	444	1	PST1_YEAST	Q12355	saccharomyc	654	6	0.8	486	1	MURE_CHLNU	Q9pk6c	chlamydia m
582	6	0.8	444	1	TBB_ACHKL	P20802	achlya kleb	655	6	0.8	486	1	NAB1_HUMAN	O13506	homo sapien
583	6	0.8	444	1	TBB_PHYCI	O59837	phytophthor	656	6	0.8	486	1	RP54_VIBAN	O08429	vibriu angu
584	6	0.8	445	1	PERG_STSTR	O56115	streptococc	657	6	0.8	487	1	GIRK_ANAPL	Q99218	anas platyr
585	6	0.8	445	1	YIEG_ECOLI	P31466	eschierichia	658	6	0.8	487	1	Y442_MYCTU	Y442_MYCTU	mycobacteri
586	6	0.8	446	1	MOC_CHICK	P01875	gallus galli	659	6	0.8	492	1	CAT1_ARARH	Q96528	arabidopsis
587	6	0.8	447	1	PWE_MEDSA	Q42920	medicago sa	660	6	0.8	492	1	CAT2_GOSHI	P13057	gossypium h
588	6	0.8	448	1	PBL5_MOUSE	Q9wvh8	mus muscullu	661	6	0.8	496	1	NIFD_ANASP	P00464	anaeana sp
589	6	0.8	448	1	PBL5_RAT	Q92908	homo sapien	662	6	0.8	492	1	CRT2_LYCES	Q9xth3	lycopersico
590	6	0.8	450	1	GAT6_HUMAN	O60151	streptococc	663	6	0.8	495	1	ACHP_RAT	P13392	rattus norv
591	6	0.8	450	1	GSHR_STSTR	P36135	saccharomyc	664	6	0.8	495	1	CSEP_COXBU	P15423	coxiiella bu
592	6	0.8	450	1	UTRH_YEAST	O32193	baecillus su	665	6	0.8	496	1	TM30_MOUSE	P15533	mus muscullu
593	6	0.8	451	1	CSSS_BACSU	P24125	lactococcus	666	6	0.8	497	1	NIFD_ANASP	P00464	anaeana sp
594	6	0.8	452	1	V51X_BPL79	Q94111	mus muscullu	667	6	0.8	499	1	C1W5_HUMAN	O95279	homo sapien
595	6	0.8	454	1	GUAD_MOUSE	Q9wt6	rattus norv	668	6	0.8	499	1	RPB2_METYA	P41558	methanococc
596	6	0.8	454	1	GUAD_RAT	O14215	schiizosach	669	6	0.8	500	1	NU4C_HORVU	O01060	hordium vul
597	6	0.8	454	1	PR11_SCHPO	Q60476	cavia porce	670	6	0.8	500	1	NU4C_ORYSA	P12127	oryza sativ
598	6	0.8	455	1	A2AC_CAVPO	P08867	citridhida f	671	6	0.8	500	1	ACHB_HUMAN	P11230	homo sapien
599	6	0.8	455	1	PGKC_CRIFA	O86564	streptomyce	672	6	0.8	501	1	ACHB_HUMAN	P11230	homo sapien
600	6	0.8	455	1	SDHL_STRCO	P71555	mycobacteri	673	6	0.8	502	1	K6B1_HUMAN	P23445	homo sapien
601	6	0.8	455	1	Y955_MYCTU	O99613	staphylococ	674	6	0.8	502	1	K6B1_RAT	P21425	rattus norv
602	6	0.8	456	1	LEU2_STAM	P58947	staphylococ	675	6	0.8	503	1	S1Z_BSPSP	P34309	bacteriophia
603	6	0.8	456	1	LEU2_STAM	P20768	staphylococ	676	6	0.8	506	1	AGR_ECOLI	P50466	eschierichia
604	6	0.8	457	1	MOC_SUNMU	P55709	rhizobium s	677	6	0.8	507	1	DHAL_MYCTU	O53743	mycobacteri
605	6	0.8	458	1	Y4YA_RHISN	P15169	homo sapien	678	6	0.8	508	1	HMCM_PIG	O02734	sus scrofa
606	6	0.8	458	1	CBPN_HUMAN	P23128	drosophila	679	6	0.8	511	1	SVS_MOUSE	P26638	mus muscullu
607	6	0.8	460	1	ME31_DROME	P33701	clostridium	680	6	0.8	511	1	YASA_ECOLI	P58229	eschierichia
608	6	0.8	460	1	GUIN2_CLOAO	O986P3	rhizobium l	681	6	0.8	511	1	YASA_SHIFL	P34183	eschierichia
609	6	0.8	460	1	NIFN_RHILC	P21592	saccharomyc	682	6	0.8	511	1	RTTI_YEAST	P81131	volvox cart
610	6	0.8	462	1	COXK_YEAST	P44587	haemophilus	683	6	0.8	512	1	PR11_YOICA	P23796	saccharomyc
611	6	0.8	462	1	YFHD_HAETN	Q948E0	klebsiella	684	6	0.8	513	1	SVS_BOVIN	P48591	homo sapien
612	6	0.8	462	1	ZRAS_KLEOX	P53453	fugu rubrip	685	6	0.8	513	1	NCAP_SENDE	P04857	sendai viru
613	6	0.8	463	1	D2DR_FUGRU	O03557	saccharomyc	686	6	0.8	517	1	NCAP_SENDE	P04858	sendai viru
614	6	0.8	464	1	GATH_YEAST	O8y649	bruceella me	687	6	0.8	517	1	VGIG_VSVVO	P04882	vesicular s
615	6	0.8	466	1	SAHH_BRIME	O92cc1	rhizobium m	688	6	0.8	517	1	VGIG_VSVVO	P04882	vesicular s
616	6	0.8	466	1	SAHH_RHIME	O27605	methanobact	689	6	0.8	517	1	VGIG_VSVVO	P04882	vesicular s
617	6	0.8	469	1	NIFD_METTH			690	6	0.8	517	1	VGIG_VSVVO	P04882	vesicular s

691	6	0.8	518	1	ATPA_MYCGE	P47641 mycoplasma	764	6	0.8	596	1	YA4C_SCHPO	Q99729 schizosacch
692	6	0.8	518	1	ATPA_MYCPN	Q50329 mycoplasma	765	6	0.8	599	1	GP63_LEICH	P15706 leishmania
693	6	0.8	521	1	SPKA_SYNY3	Q9fab3 synchocyst	766	6	0.8	600	1	OPPA_LACL	Q07741 lactococcus
694	6	0.8	524	1	FIXG_RHIME	P18396 rhizobium m	767	6	0.8	600	1	OPPA_LACL	Q09144 lactococcus
695	6	0.8	524	1	NCAP_PILHW	P36354 human parai	768	6	0.8	602	1	GP63_LEIMA	P08148 leishmania
696	6	0.8	524	1	NCAP_PILHW	P26590 human parai	769	6	0.8	603	1	RPB1_METTH	Q27124 methanobact
697	6	0.8	524	1	NCAP_SENDS	P27563 sendai viru	770	6	0.8	604	1	RPB1_METTW	P09845 methanobact
698	6	0.8	524	1	NCAP_SENDF	Q07097 sendai viru	771	6	0.8	604	1	XJGL_ECOLI	P39336 escherichia
699	6	0.8	525	1	AROF_ARATH	P48778 pan troglod	772	6	0.8	605	1	GLMS_CHLMU	Q9pl4 c glucosami
700	6	0.8	525	1	GOR_PANTR	P48778 pan troglod	773	6	0.8	605	1	GLMS_CHLTR	O84823 c glucosami
701	6	0.8	528	1	YJCC_ECOLI	P32701 escherichia	774	6	0.8	607	1	DPOL_BPMW2	O64235 mycobacteri
702	6	0.8	528	1	YCR2_ERWHE	Q01336 erwinia her	775	6	0.8	607	1	LEUL_MYCLE	O9cb76 mycobacteri
703	6	0.8	529	1	YCH1_CAEEL	P36338 caenorhabdi	776	6	0.8	607	1	PESC_SCHPO	O60164 schizosacch
704	6	0.8	529	1	YPC1_CAEEL	Q11178 caenorhabdi	777	6	0.8	608	1	YD6C_SCHPO	Q10324 schizosacch
705	6	0.8	531	1	Y372_AQUAE	O66695 aquifex aeo	778	6	0.8	609	1	GLMS_HAEIN	P44708 h glucosami
706	6	0.8	532	1	SPG7_DICDI	P22698 dictyosteli	779	6	0.8	613	1	ENP4_MOUSE	Q9dbt4 mus musculu
707	6	0.8	533	1	UD12_MOUSE	P70691 mus musculu	780	6	0.8	614	1	RPB1_METVA	P11557 methanococ
708	6	0.8	534	1	P4H1_HUMAN	P13674 homo sapien	781	6	0.8	614	1	Y268_CHLMU	Q9pl40 chlamydia m
709	6	0.8	535	1	ARSB_FELCA	P33727 felis siive	782	6	0.8	616	1	ENP4_HUMAN	Q9y227 homo sapien
710	6	0.8	539	1	M24_STRPY	P13379 streptococc	783	6	0.8	616	1	HTPG_BORBU	P42555 borrelia bu
711	6	0.8	540	1	HUF2_CHLKE	Q39524 chlorella k	784	6	0.8	620	1	TERM_ADBB2	O55438 bovine aden
712	6	0.8	541	1	CDY2_HUMAN	Q9y6f7 homo sapien	785	6	0.8	623	1	YD54_MYCTU	P11024 mycobacteri
713	6	0.8	542	1	NCAP_P12HT	P21737 human parai	786	6	0.8	625	1	AMYG_NEUCR	P14804 neurospora
714	6	0.8	542	1	PGMU_AGRUT	P39671 agrobacteri	787	6	0.8	625	1	RSC4_YEAST	Q22206 saccharomyc
715	6	0.8	543	1	FAT2_YEAST	P38137 saccharomyc	788	6	0.8	626	1	GIDA_CIOAB	Q97cw3 clostidium
716	6	0.8	544	1	CH63_RHIME	Q930y0 rhizobium m	789	6	0.8	627	1	UVRC_ANASP	Q9ytc19 anabaena sp
717	6	0.8	545	1	CH63_BRAJA	P15932 salmonella	790	6	0.8	628	1	LU_HUMAN	P50895 homo sapien
718	6	0.8	546	1	CHOD_STRSQ	P12676 streptomyce	791	6	0.8	630	1	ESAB_TRYBB	P23799 trypanosoma
719	6	0.8	546	1	YTE4_CAEEL	Q17865 caenorhabdi	792	6	0.8	630	1	ESAB_TRYBB	P26337 trypanosoma
720	6	0.8	549	1	CH62_BRAJA	P35861 bradyrhizob	793	6	0.8	633	1	GIAL_THETN	Q8ar68 thermoaner
721	6	0.8	550	1	FLIC_SHIFL	Q08860 shigella fl	794	6	0.8	633	1	GIA2_THETN	Q8r6k9 thermoaner
722	6	0.8	551	1	AMTB_THETU	P19584 thermoaner	795	6	0.8	633	1	Y561_HAEIN	P44016 haemophilus
723	6	0.8	551	1	FLGK_SALTY	P15932 salmonella	796	6	0.8	634	1	YKCA_CAEEL	P42083 caenorhabdi
724	6	0.8	552	1	HIS5_YEAST	P33734 saccharomyc	797	6	0.8	636	1	RPB1_METJA	Q60181 methanococ
725	6	0.8	552	1	LON2_BACSU	P42425 bacillus su	798	6	0.8	638	1	YHES_HAEIN	P44808 haemophilus
726	6	0.8	553	1	HIS5_EMENT	Q9p4D9 emericella	799	6	0.8	646	1	GP63_LEIME	P43150 leishmania
727	6	0.8	554	1	MAS5_YEAST	P30952 saccharomyc	800	6	0.8	647	1	NTP1_MSEPV	Q9yws9 melanoplus
728	6	0.8	554	1	MAS2_YEAST	P21826 saccharomyc	801	6	0.8	649	1	METX_YEAST	Q04533 saccharomyc
729	6	0.8	554	1	VL2_HPV41	P27558 human papil	802	6	0.8	650	1	DD17_HUMAN	Q92841 homo sapien
730	6	0.8	557	1	ILVD_AERPE	Q9y988 aeropyrum p	803	6	0.8	651	1	YHJK_ECOLI	P37649 escherichia
731	6	0.8	557	1	PKS3_BACSU	P40806 bacillus su	804	6	0.8	652	1	FAS1_DROME	P10674 drosophila
732	6	0.8	557	1	SVQ_HAEIN	P43831 haemophilus	805	6	0.8	655	1	DNK1_SYNP7	P50020 synchococ
733	6	0.8	558	1	CSG_METJA	Q58232 methanococ	806	6	0.8	655	1	NDC1_YEAST	P32500 saccharomyc
734	6	0.8	559	1	FTHS_MOOTH	P21164 moorella th	807	6	0.8	656	1	V091_FOMPV	O72896 fowlpox vir
735	6	0.8	559	1	SVQ_PASMU	P57847 pasteurella	808	6	0.8	656	1	YK79_MYCTU	Q10687 mycobacteri
736	6	0.8	560	1	INRI_BOVIN	Q04790 bos taurus	809	6	0.8	657	1	MTHR_YEAST	P46151 saccharomyc
737	6	0.8	561	1	TREC_BACSU	P39795 bacillus su	810	6	0.8	658	1	ABD3_RAT	P16970 rattus norv
738	6	0.8	564	1	YK62_AQUAE	O67301 aquifex aeo	811	6	0.8	659	1	VST2_HEVME	Q03500 hepatitis e
739	6	0.8	566	1	PKNE_MYCTU	P72001 mycobacteri	812	6	0.8	661	1	LOX1_HUMAN	P16050 homo sapien
740	6	0.8	566	1	TS13_MOUSE	Q01755 mus musculu	813	6	0.8	662	1	FAS1_SCHAM	P10675 schistocerc
741	6	0.8	569	1	MCRA_METBA	P07962 methanosarc	814	6	0.8	662	1	TRB2_RHISN	P55399 rhizobium s
742	6	0.8	569	1	YAP3_YEAST	P32329 saccharomyc	815	6	0.8	663	1	GRIA_BACCE	O85467 bacillus ce
743	6	0.8	569	1	Y599_MYCPN	P75324 mycoplasma	816	6	0.8	667	1	EM70_YEAST	P18594 agrobacteri
744	6	0.8	570	1	UL49_HCMVA	P16786 human cytom	817	6	0.8	668	1	VIDA_AGR5	P32802 saccharomyc
745	6	0.8	574	1	C4GP_DROME	Q9vyt4 drosophila	818	6	0.8	669	1	DNL3_SYNY3	P72588 synchocyst
746	6	0.8	574	1	MAO1_ECOLI	P26616 escherichia	819	6	0.8	671	1	VINE_HUMAN	P13464 agrobacteri
747	6	0.8	574	1	ME22_SCHPO	O60170 schizosacch	820	6	0.8	671	1	VID4_AGRRH	Q60504 homo sapien
748	6	0.8	575	1	TERM_ADEG1	Q64752 avian adeno	821	6	0.8	673	1	KNS2_HUMAN	Q9bw19 homo sapien
749	6	0.8	576	1	PABP_YEAST	P04147 saccharomyc	822	6	0.8	674	1	MUTL_CLOPE	O8bw18 clostridium
750	6	0.8	576	1	PEX5_PICPA	P33292 pichia past	823	6	0.8	676	1	HMUR_YERPE	Q56989 yersinia pe
751	6	0.8	576	1	REC5_BACSU	P17894 bacillus su	824	6	0.8	678	1	CICL_RABIT	P51804 oryctolagus
752	6	0.8	578	1	YTFM_HAEIN	P44038 haemophilus	825	6	0.8	680	1	KALM_HUMAN	P23352 homo sapien
753	6	0.8	579	1	LDL1_KLUJA	Q12627 kluyveromyc	826	6	0.8	682	1	CY2G_RAT	P22717 rattus norv
754	6	0.8	583	1	API_KLUJA	P56095 kluyveromyc	827	6	0.8	687	1	AKA8_RAT	Q63014 rattus norv
755	6	0.8	584	1	MKS1_YEAST	P34072 saccharomyc	828	6	0.8	687	1	CICK_RABIT	P51803 oryctolagus
756	6	0.8	585	1	PTRR_DIDNA	P35107 didelphis m	829	6	0.8	687	1	HEMR_YEREN	P31498 yersinia en
757	6	0.8	590	1	GF63_LEIDO	P23223 leishmania	830	6	0.8	687	1	INA_BACTL	P23382 bacillus th
758	6	0.8	593	1	GRN_HUMAN	P28799 h granulins	831	6	0.8	688	1	VID4_AGRTU	P09817 agrobacteri
759	6	0.8	594	1	DCP2_SCHPO	Q92345 schizosacch	832	6	0.8	689	1	YBP3_YEAST	P38227 saccharomyc
760	6	0.8	594	1	RCO3_NEUCR	Q92253 neurospora	833	6	0.8	690	1	HELS_METTH	O26901 methanobact
761	6	0.8	595	1	ACR2_NEUCR	P78704 neurospora	834	6	0.8	690	1	PERO_DROME	Q01603 drosophila
762	6	0.8	595	1	DPOL_BPML5	Q05254 mycobacteri	835	6	0.8	691	1	OAT5_HUMAN	Q9y616 homo sapien
763	6	0.8	595	1	YE72_HUMAN	Q9nx95 homo sapien	836	6	0.8	692	1	POL_SOCNV	P15629 soybean chl

Q99729 schizosacch	Q15706 leishmania	Q07741 lactococcus	Q09144 lactococcus	P08148 leishmania	Q27124 methanobact	P09845 methanobact	P39336 escherichia	Q9pl4 c glucosami	O84823 c glucosami	O64235 mycobacteri	O9cb76 mycobacteri	O60164 schizosacch	Q10324 schizosacch	P44708 h glucosami	Q9dbt4 mus musculu	P11557 methanococ	Q9pl40 chlamydia m	Q9y227 homo sapien	P42555 borrelia bu	O55438 bovine aden	P23799 trypanosoma	P26337 trypanosoma	Q8ar68 thermoaner	Q8r6k9 thermoaner	P44016 haemophilus	P42083 caenorhabdi	Q60181 methanococ	P44808 haemophilus	P43150 leishmania	Q9yws9 melanoplus	Q04533 saccharomyc	Q92841 homo sapien	P37649 escherichia	P10674 drosophila	P50020 synchococ	P32500 saccharomyc	O72896 fowlpox vir	Q10687 mycobacteri	P46151 saccharomyc	P16970 rattus norv	Q03500 hepatitis e	P16050 homo sapien	P10675 schistocerc	P55399 rhizobium s	O85467 bacillus ce	P18594 agrobacteri	P32802 saccharomyc	P72588 synchocyst	P13464 agrobacteri	Q60504 homo sapien	Q9bw19 homo sapien	O8bw18 clostridium	Q56989 yersinia pe	P51804 oryctolagus	P23352 homo sapien	P22717 rattus norv	Q63014 rattus norv	P51803 oryctolagus	P31498 yersinia en	P23382 bacillus th	P09817 agrobacteri	P38227 saccharomyc	O26901 methanobact	Q01603 drosophila	Q9y616 homo sapien	P15629 soybean chl
--------------------	-------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	-------------------	--------------------	--------------------	-------------------	--------------------	-------------------	-------------------	--------------------	--------------------	--------------------	-------------------	------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------

837	6	0.8	593	1	THPA_HUMAN	P42166	homo sapien	910	6	0.8	821	1	PSGA_PYRHO	O57830	pyrococcus
838	6	0.8	701	1	LOXR_HUMAN	O75342	homo sapien	911	6	0.8	822	1	NFC1_PIG	O77638	sus scrofa
839	6	0.8	702	1	SYT_MYCL	O07151	mycobacteri	912	6	0.8	823	1	NSP1_YEAST	P14907	saccharomyc
840	6	0.8	709	1	SYT_ARATH	O04630	arabidopsis	913	6	0.8	826	1	CRA_BACCH	O96597	bacillus th
841	6	0.8	710	1	GSOD_ERWCH	O01565	erwinia chr	914	6	0.8	826	1	PRUA_PSESP	O08017	pseudomonas
842	6	0.8	716	1	IKKE_HUMAN	O14164	homo sapien	915	6	0.8	826	1	RIR1_EBV	P03190	epstein-bar
843	6	0.8	716	1	RBP1_DHVIL	P27153	dhori virus	916	6	0.8	829	1	RREW_YEAST	P25328	saccharomyc
844	6	0.8	717	1	IKKE_MOUSE	O91018	mus musculu	917	6	0.8	836	1	CEG1_HUMAN	O99490	homo sapien
845	6	0.8	719	1	FLBA_EMENT	P38093	emeritella	918	6	0.8	836	1	TGML_RABIT	P22758	oryctolagus
846	6	0.8	720	1	PSAB_EPHW	O9muj8	ephedra twe	919	6	0.8	837	1	SECA_LISIN	O92773	liesteria in
847	6	0.8	724	1	PCRA_BACST	P56255	bacillus st	920	6	0.8	837	1	SECA_LISMO	P47847	liesteria mo
848	6	0.8	730	1	PIOD_CABEL	O20679	caenorhabdi	921	6	0.8	837	1	SM4G_MOUSE	O9wh7	mus musculu
849	6	0.8	732	1	ATZN_ECOLI	P37617	escherichia	922	6	0.8	838	1	PHS3_RAT	P53534	rattus norv
850	6	0.8	732	1	PRIA_ECOLI	P17888	escherichia	923	6	0.8	838	1	SM4G_HUMAN	O9tntn	homo sapien
851	6	0.8	734	1	PSAB_ARATH	P56767	arabidopsis	924	6	0.8	840	1	UREA_CANEN	P07374	canavalia e
852	6	0.8	734	1	PSAB_ORYSA	P12152	oryza sativ	925	6	0.8	841	1	SECA_BACCU	P28366	bacillus su
853	6	0.8	734	1	YKR2_CABEL	P34308	caenorhabdi	926	6	0.8	842	1	AXN_XENLA	O9y9y0	xenopus lae
854	6	0.8	735	1	PSAB_MAIZE	P04967	zea mays (m	927	6	0.8	842	1	PHS2_BOVIN	P79334	bos taurus
855	6	0.8	737	1	PEN_DROME	O61345	drosophila	928	6	0.8	842	1	PHS2_HUMAN	P11217	homo sapien
856	6	0.8	738	1	TRFM_HUMAN	P08582	homo sapien	929	6	0.8	842	1	PHS2_MOUSE	O9wub3	mus musculu
857	6	0.8	740	1	PECI_PIG	O95242	sus scrofa	930	6	0.8	842	1	PHS2_RABIT	P00489	oryctolagus
858	6	0.8	741	1	CATA_ARCFU	O28050	archaeoglob	931	6	0.8	842	1	PHS2_RAT	P09812	rattus norv
859	6	0.8	747	1	VVIC_BPT7	P03725	bacteriopia	932	6	0.8	842	1	PHS2_SHEEP	O18751	ovis aries
860	6	0.8	750	1	YK57_YEAST	P34231	saccharomyc	933	6	0.8	843	1	PHS3_HUMAN	P11216	homo sapien
861	6	0.8	752	1	CUPE_STREP	P35594	streptococc	934	6	0.8	843	1	SECA_STAUV	O06446	staphylococ
862	6	0.8	755	1	ADIA_ECOLI	P28629	escherichia	935	6	0.8	844	1	APB2_YEAST	P13454	saccharomyc
863	6	0.8	758	1	PURL_CORAM	O9rtw9	corynebacte	936	6	0.8	844	1	PHSG_DROME	O9xtl9	drosophila
864	6	0.8	759	1	ARY1_CALVI	P28513	calliphora	937	6	0.8	844	1	SECA_STACA	P47994	staphylococ
865	6	0.8	759	1	YEN1_YEAST	P40028	saccharomyc	938	6	0.8	847	1	PHS1_HUMAN	P06737	homo sapien
866	6	0.8	763	1	TSHR_BOVIN	O27987	bos taurus	939	6	0.8	850	1	PHS1_MOUSE	O9ae01	mus musculu
867	6	0.8	764	1	TSHR_CANFA	P14763	canis famli	940	6	0.8	850	1	PHS1_RAT	P08901	rattus norv
868	6	0.8	764	1	TSHR_HUMAN	P16473	homo sapien	941	6	0.8	854	1	VGIB_RHOM6	P89053	rhesus cyto
869	6	0.8	764	1	TSHR_MOUSE	P47750	mus musculu	942	6	0.8	854	1	XR52_YEAST	P33301	saccharomyc
870	6	0.8	764	1	TSHR_RAT	P21463	rattus norv	943	6	0.8	859	1	YD48_MYCTU	O11018	mycobacteri
871	6	0.8	765	1	TSHR_SHEEP	P56495	ovis aries	944	6	0.8	862	1	112S_HUMAN	O99665	homo sapien
872	6	0.8	765	1	Y008_HUMAN	O15398	homo sapien	945	6	0.8	865	1	PMU1_HUMAN	O43490	homo sapien
873	6	0.8	766	1	BCSB_SALTI	O82290	salmomella	946	6	0.8	865	1	SECA_HELPJ	O92157	helicobacte
874	6	0.8	766	1	BCSB_SALTY	O931n1	salmomella	947	6	0.8	865	1	SECA_HELPJ	O25475	helicobacte
875	6	0.8	768	1	TRPG_EMENT	P06531	emeritella	948	6	0.8	868	1	MCW2_YEAST	P29469	saccharomyc
876	6	0.8	768	1	YDDB_ECOLI	P25907	escherichia	949	6	0.8	868	1	NR80_YEAST	P33420	saccharomyc
877	6	0.8	771	1	PCRA_MYCTU	P71561	mycobacteri	950	6	0.8	869	1	CRAC_ECOLI	P25733	escherichia
878	6	0.8	776	1	KUPI_CHLRE	P46870	chlamydomon	951	6	0.8	872	1	DEOI_SULOH	O50607	sulfurispina
879	6	0.8	778	1	SEA2_MYCL	O32922	mycobacteri	952	6	0.8	874	1	ATCL_MYCGB	P47317	mycoplasma
880	6	0.8	779	1	CA11_BOVIN	P02453	bos taurus	953	6	0.8	875	1	SECA_BUCAI	P57297	buhnera ap
881	6	0.8	782	1	CHAO_TRICA	P82963	tribolium c	954	6	0.8	877	1	SECA_GUINR	O78441	guillardia s
882	6	0.8	782	1	ZAM_SYNY3	O46363	synecocyst	955	6	0.8	878	1	SECA_ANTSP	O05461	antithamio
883	6	0.8	784	1	TIAR2_HUMAN	O60603	homo sapien	956	6	0.8	884	1	SECA_OLITU	O32743	olisthodisc
884	6	0.8	784	1	TIAR2_NACFA	O95053	macaca faec	957	6	0.8	884	1	SECA_PORPU	P51381	porphyra pu
885	6	0.8	785	1	IF16_HUMAN	O16666	homo sapien	958	6	0.8	886	1	OBPI_HAEIN	P45119	haemophilus
886	6	0.8	787	1	PFL_LACLA	O32797	lactococcus	959	6	0.8	887	1	OBP_HEVER	P28947	equine herp
887	6	0.8	787	1	PFL_LACLC	O32799	lactococcus	960	6	0.8	888	1	SECA_ODOST	P46649	odontella s
888	6	0.8	787	1	XPKA_LACPE	O937f6	lactobacill	961	6	0.8	888	1	SYA_FRILLO	O9nng5	rhizobium l
889	6	0.8	790	1	DNLI_ARATH	O42572	arabidopsis	962	6	0.8	890	1	YOIN_ECOLI	P39838	escherichia
890	6	0.8	790	1	SUNI_MOUSE	O9d666	mus musculu	963	6	0.8	891	1	SECA_PAVLU	O01570	pavlova lut
891	6	0.8	790	1	SYFB_CHLMO	O9pjr8	chlamydia m	964	6	0.8	899	1	SECA_BORBU	O07497	borrella bu
892	6	0.8	790	1	SYFB_CHLTR	O84481	chlamydia t	965	6	0.8	900	1	YB75_YEAST	P38321	saccharomyc
893	6	0.8	792	1	YB5K_SSV1	P20210	sulfolobus	966	6	0.8	901	1	A180_MOUSE	O61548	mus musculu
894	6	0.8	795	1	Y210_HUMAN	O92609	homo sapien	967	6	0.8	901	1	CH14_MAIZE	O24585	zea mays (m
895	6	0.8	796	1	DHG_ECOLI	P15877	escherichia	968	6	0.8	901	1	DSOC2_HUMAN	O02487	bos taurus
896	6	0.8	796	1	PHK_CLOAB	O97f63	clostridium	969	6	0.8	901	1	SECA_ECOLI	P10408	escherichia
897	6	0.8	797	1	SACB_STRMU	P11701	streptococc	970	6	0.8	901	1	SECA_HAEIN	P43803	haemophilus
898	6	0.8	798	1	YFR2_MOUSE	O9jxk3	mus musculu	971	6	0.8	902	1	NRFC4_HUMAN	O14934	homo sapien
899	6	0.8	798	1	YCOO_MYCPN	O50088	mycoplasma	972	6	0.8	904	1	SECA_RHCCA	P51966	rhodobacter
900	6	0.8	806	1	SECA_MYCGB	P47318	mycoplasma	973	6	0.8	906	1	SECA_RICR	O9cxc7	ricketsia
901	6	0.8	808	1	SEA2_MYCTU	O50612	mycobacteri	974	6	0.8	912	1	PGCB_BOVIN	O28062	bos taurus
902	6	0.8	808	1	SECA_MYCPN	P75559	mycoplasma	975	6	0.8	915	1	A180_RAT	O05140	rattus norv
903	6	0.8	809	1	TORZ_ECOLI	P46923	escherichia	976	6	0.8	916	1	SECA_TREPA	O08339	treponema p
904	6	0.8	810	1	TIRA_ECOLI	O07736	escherichia	977	6	0.8	919	1	AWY_STRLI	O05884	streptomyce
905	6	0.8	812	1	FAED_ECOLI	P06970	escherichia	978	6	0.8	919	1	DNLI_HUMAN	P18858	homo sapien
906	6	0.8	812	1	PSAB_PYRRTU	P42850	pyrococcus	979	6	0.8	919	1	NOS3_RAT	O62600	rattus norv
907	6	0.8	819	1	PSAB_PYRAB	O92zh7	pyrococcus	980	6	0.8	919	1	RPO2_CAPVK	P16716	capripoxvir
908	6	0.8	820	1	GLGB_ORYSA	O01401	oryza sativ	981	6	0.8	922	1	YKFO_YEAST	P35736	saccharomyc
909	6	0.8	820	1	TRBE_AGRITU	P54910	agrobacteri	982	6	0.8	923	1	SECA_CAUCR	P38380	caulobacter

983 6 0.8 924 1 AACT DROME
984 6 0.8 926 1 SC24 YEAST
985 6 0.8 929 1 SYA HALNI
986 6 0.8 932 1 SECA SYN3
987 6 0.8 933 1 A8B2 HUMAN
988 6 0.8 934 1 Y321 MYCCE
989 6 0.8 937 1 ODOI BACSU
990 6 0.8 940 1 SEAI MYCLE
991 6 0.8 940 1 SECA STRGR
992 6 0.8 943 1 DIP2 YEAST
993 6 0.8 945 1 A8B3 HUMAN
994 6 0.8 946 1 CITC YEAST
995 6 0.8 947 1 SECA STRCO
996 6 0.8 948 1 PMA5 ARATH
997 6 0.8 948 1 SECA MYCBO
998 6 0.8 948 1 SECA SYN7
999 6 0.8 949 1 SEAI MYCTU
1000 6 0.8 956 1 NUT1 MAGGR

ALIGNMENTS

RESULT 1
DXS_BACHD STANDARD; PRT; 629 AA.
AC Q9K371;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 4.1.3.37) (1-
DXS OR BH2779.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RC MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Catalyzes the acyloln condensation reaction between C
atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).
CC -!- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
deoxy-D-xylulose 5-phosphate + CO(2).
CC -!- COFACTOR: Thiamine pyrophosphate (By similarity).
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
CC -!- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY. DXS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001516; BAB06498.1; -
DR InterPro; IPR000360; Transketolase.
DR Pfam; PF02779; transket_pyr; 1.
DR Pfam; PF02780; transketolase_C; 1.
DR TIGRFAMs; TIGR00204; dxs; 1.
DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.

KW Lyase; Flavoprotein; Thiamine pyrophosphate;
KW Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
SQ SEQUENCE 629 AA; 69279 MW; 7B5909B2F427C11E CRC64;
Query Match 1.2%; Score 9; DB 1; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.79; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
Qy 470 GHYKIESGE 478
| | | | | | | | | |
Db 299 GHYKIESGE 307
| | | | | | | | | |
RESULT 2
RNH2_PASMU STANDARD; PRT; 197 AA.
AC P57986;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease HII (EC 3.1.26.4) (RNase HII).
GN RNHB OR PM1998.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: THIS ENZYME IS AN ENDONUCLEASE THAT DEGRADES THE RNA OF
RNA-DNA HYBRIDS SPECIFICALLY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
phosphomononester.
CC -!- COFACTOR: MANGANESE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE RNASE HII FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006236; AAK04082.1; -
DR HSSP; Q57599; 1EKE.
DR InterPro; IPR001352; RNase HII/HIII.
DR Pfam; PF01351; RNase HII; 1.
KW Hydrolase; Nuclease; Endonuclease; Manganese; Complete proteome.
FT ACT SITE 16 BY SIMILARITY.
FT ACT SITE 16 108 BY SIMILARITY.
FT ACT SITE 126 126 BY SIMILARITY.
SQ SEQUENCE 197 AA; 21589 MW; CAA9F9C6442C7BB CRC64;
Query Match 1.1%; Score 8; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 360 EIKQKALA 367
| | | | | | | | | |
Db 60 EIKQKALA 67
| | | | | | | | | |
RESULT 3
ICPO_HSVB STANDARD; PRT; 532 AA.
ID ICPO_HSVB
AC P28930;
DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trans-acting transcriptional protein ICPO.
 GN 63.
 OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
 OC Viruses; deDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirinae.
 NCBI_TaxID=31520;
 RX SEQUENCE FROM N.A.
 RX MEDLINE=92295566; PubMed=1318606;
 RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
 RT "The DNA sequence of equine herpesvirus-1";
 RL Virology 189:304-316(1992).
 RN [2]
 RP STRUCTURE BY NMR OF 1-63.
 RX MEDLINE=94087718; PubMed=8263911;
 RA Everett R.D., Barlow P.N., Milner A., Luisi B., Orr A., Hope G.,
 RA Lyon D.;
 RT "A novel arrangement of zinc-binding residues and secondary structure
 in the C3HC4 motif of an alpha herpes virus protein family.";
 RL J. Mol. Biol. 234:1038-1047(1993).
 RN [3]
 RP STRUCTURE BY NMR OF 1-63.
 RX MEDLINE=94172642; PubMed=8126734;
 RA Barlow P.N., Luisi B., Milner A., Elliott M., Everett R.D.;
 RT "Structure of the C3HC4 domain by 1H-nuclear magnetic resonance
 spectroscopy. A new structural class of zinc-finger.";
 RL J. Mol. Biol. 237:201-211(1994).
 CC -1- SIMILARITY: CONTRAINS 1 RING-TYPE ZINC FINGER.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICPO PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M86664; AAB02498.1; -
 DR PIR: 136801; WZBER5.
 DR PDB: 1CHC; 30-APR-94.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF00097; zf-C3HC4.1.
 DR SMART: SM00184; RING.1.
 DR PROSITE: PS00518; ZF_RING_1.1.
 DR PROSITE: PS50089; ZF_RING_2.1.
 KM Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
 KM DNA-binding; 3D-structure.
 FT ZN FING 8 47
 FT METAL 8 47 RING-TYPE.
 FT METAL 11 11 ZINC 1.
 FT METAL 24 24 ZINC 2.
 FT METAL 26 26 ZINC 2.
 FT METAL 29 29 ZINC 2.
 FT METAL 32 32 ZINC 1.
 FT METAL 43 43 ZINC 2.
 FT METAL 46 46 ZINC 2.
 FT DOMAIN 210 217 POLY-SER.
 FT STRAND 19 21
 FT TURN 22 25
 FT STRAND 26 28
 FT TURN 30 31
 FT HELIX 32 39
 FT TURN 43 46
 FT TURN 46 54
 SQ SEQUENCE 532 AA; 58629 MW; B4CB7E16FA26FDFA CRC64;
 Query Match 1.1%; Score 8; DB 1; Length 532;
 Best Local Similarity 100.0%; Pred. No. 7.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 199 ADVSGSLN 206
 DB 323 ADVSGSLN 330
 RESULT 4
 ID LEU1_MYCTU STANDARD; PRT; 644 AA.
 AC P96420; O69677;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 2-Isopropylmalate synthase (EC 4.1.3.12) (Alpha-isopropylmalate
 synthase) (Alpha-IPM synthetase).
 GN LEU1 OR RV3710 OR MT3813 OR MTV025.058.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 RX SEQUENCE FROM N.A.
 RX STRAIN=H37Rv;
 RA Chancham W., Namwat W., Palittapongprap P.;
 RT "Putative Mycobacterium tuberculosis alpha-isopropyl malate
 synthase gene."
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=H37Rv;
 RX MEDLINE=96295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Uegels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
 RA Sultun J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=CDC 1551 / Oshkosh;
 RA Pleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains".
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: Catalyzes the condensation of the acetyl group of
 CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
 CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
 CC -1- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA =
 CC acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.
 CC -1- PATHWAY: Leucine biosynthesis; first step.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
 CC SYNTHASE FAMILY. LEU1A 2 SUBFAMILY.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 49 and 77.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U88526; AAB48096.1; ALT_FRAME.

```

DR EMBL; AL022121; CAA18032.1; -.
DR EMBL; AE007178; AAK48181.1; ALT_INIT.
DR TIGR; MT3813; -.
DR TubercuList; RV3710; -.
DR InterPro; IPR002034; AIPM/Hcit_synth.
DR InterPro; IPR000891; HMGL-like.
DR Pfam; PF00682; HMGL-like; 1.
DR TIGRFAMs; TIGR00970; leuA_yeast; 1.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
KW Leucine biosynthesis; Lyase; Complete proteome.
FT DOMAIN 577 614 INSERT.
FT CONFLICT 628 633 TTASUR -> NRPA (IN REF. 1).
FT CONFLICT 642 644 AAR -> RHARTALN (IN REF. 1).
SQ SEQUENCE 644 AA; 70113 MW; 881A9AAEA7F8FB71 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 644;
Best Local Similarity 100.0%; Pred. No. 8.8; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

QY 585 GVETBISG 592
DB 522 GVETBISG 529
|||||

RESULT 5
FOXA_SALTY
ID FOXA_SALTY STANDARD; PRT; 702 AA.
AC Q56145.
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferrioxamine B receptor precursor.
GN FOXA OR STM0364.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
[1] SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A. / ATCC 700720;
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterson R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
[2] SEQUENCE OF 1-374 FROM N.A.
RP SEQUENCE OF 1-374 FROM N.A.
RC STRAIN=SL1344;
RX MEDLINE=95203118; PubMed=10103258;
RA Kingsley R.A., Reissbrodt R., Rabsch W., Ketley J.M., Tsolis R.M.,
Everest P., Dougan G., Baumeister A.J., Roberts M., Williams P.H.;
RA "Ferrioxamine-mediated iron(III) utilization by Salmonella
enterica.";
RT Appl. Environ. Microbiol. 65:1610-1618(1999).
[3] SEQUENCE OF 1-244 FROM N.A.
RP SEQUENCE OF 1-244 FROM N.A.
RC STRAIN=ATCC 14028;
RX MEDLINE=95370139; PubMed=7642488;
RA Tsolis R.M., Baunler A.J., Stojiljkovic I., Heffron F.;
RA "Fur regulon of Salmonella typhimurium: identification of new iron-
regulated genes.";
RT J. Bacteriol. 177:4628-4637(1995).
RL J. Bacteriol. 177:4628-4637(1995).
CC -!- FUNCTION: FERROXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE
TONB PROTEIN (BY SIMILARITY). MAY PLAY A ROLE IN INTESTINAL
COLONIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008712; AAL19318.1; -.
DR EMBL; AF060876; AAC15464.1; -.
DR EMBL; U62282; AAB04552.1; -.
DR HSSP; P06971; 1BY5.
DR StyGene; SG10646; foxA.
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW Transport; TonB box; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 702 FERROXAMINE B RECEPTOR.
FT CONFLICT 246 249 MAYP -> SPYL (IN REF. 2).
FT CONFLICT 255 255 G -> R (IN REF. 2).
FT CONFLICT 279 279 F -> S (IN REF. 2).
FT CONFLICT 352 352 A -> P (IN REF. 2).
FT CONFLICT 368 374 HRLLVGI -> SSSAGGD (IN REF. 2).
SQ SEQUENCE 702 AA; 77685 MW; 7E2CB405B09E860D CRC64;

Query Match 1.1%; Score 8; DB 1; Length 702;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 POSVSIIT 81
DB 60 POSVSIIT 67
|||||

RESULT 6
FPTA_PSEAE
ID FPTA_PSEAE STANDARD; PRT; 720 AA.
AC P42512.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fe(III)-pyochelin receptor precursor.
GN FPTA OR PA4221.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
[1] SEQUENCE FROM N.A., AND SEQUENCE OF 39-51.
RP SEQUENCE FROM N.A. / IA602;
RC STRAIN=PAO / IA602;
RX MEDLINE=94117363; PubMed=8288523;
RA Ankenbauer R.G.; Quan H.N.;
RT "Fpta, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a
phenolate siderophore receptor homologous to hydroxamate siderophore
receptors.";
RT J. Bacteriol. 176:307-319(1994).
[2] SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltry L., Tolentino E., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: HIGH-AFFINITY OUTER MEMBRANE RECEPTOR REQUIRED FOR THE
TRANSPORT OF FE(III)-PYOCHELIN.

```



```

CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03161; AAC43213.1; -
DR EMBL; AE004839; AAG07609.1; -
DR InterPro; IPR000531; TonB_BoxC.
DR Pfam; PF00593; TonB_BoxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR Outer membrane; Iron transport; Transport; Signal; Receptor;
KM Complete proteome.
FT SIGNAL 1 38
FT CHAIN 39 720 FE(III)-PYOCHELIN RECEPTOR.
FT SITE 703 720 TONB C-TERMINAL BOX.
SQ SEQUENCE 720 AA; 79992 MW; DA796313116B0CC2 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 720;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 VNLVRRKP 177
Db 176 VNLVRRKP 183

RESULT 7
FPVA_PSEAE STANDARD; PRT; 815 AA.
AC P48632; Q51339;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferripyoverdine receptor precursor.
GN FPVA OR PA2398.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-64.
RC STRAIN=CD10;
RA MEDLINE=93328663; PubMed=8335619;
RA Poole K., Neshat S., Krebs K., Heinrichs D.E.;
RT "Cloning and nucleotide sequence analysis of the ferripyoverdine
RT receptor gene fpva of Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:4597-4604(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RA Lamont I.L.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RA MEDLINE=20437337; PubMed=10984043;
RA Hickey C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Goltzy U., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lapidis K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiser J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: RECEPTOR FOR THE SIDEROPHORE, FERRIPYOVERDINE.

```

```

CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY THE SIDEROPHORE, PYOVERDINE, AND UNDER IRON
CC STARVATION CONDITIONS.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-16 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10210; AAA25819.1; -
DR EMBL; U07359; AAB60199.1; -
DR EMBL; AE004666; AAG05786.1; -
DR PIR; A40601; A40601.
DR InterPro; IPR000531; TonB_BoxC.
DR Pfam; PF00593; TonB_BoxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR Outer membrane; Iron transport; Transport; TonB_Box; Signal; Receptor;
KM Complete proteome.
FT SIGNAL 1 43
FT CHAIN 44 815 FERRIPYOVERDINE RECEPTOR.
FT SITE 798 815 TONB C-TERMINAL BOX.
FT CONFLICT 716 716 Y -> F (IN REF. 1).
FT CONFLICT 745 746 MISSING (IN REF. 1).
FT CONFLICT 750 750 S -> R (IN REF. 1).
SQ SEQUENCE 815 AA; 91167 MW; A11AE4AA5A290F35 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 815;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 495 GAVYDLND 502
Db 574 GAVYDLND 581

RESULT 8
THIO_CHLMU STANDARD; PRT; 102 AA.
ID THIO_CHLMU
AC Q9PJF3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thioiodoxin (TRX).
GN TRXA OR TC0826.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mopn / Ni99;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uetzelback T., Berry K., Bass S.,
RA Linner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC -----
 CC EMBL; AE002349; AAF39627.1; --
 DR HSP; F80579; IQW.
 DR TIGR; TC0826; --
 DR InterPro; IPR000063; ThioRed.
 DR Pfam; PF00085; thioRed; 1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR TIGRPFAMs; TIGR01068; thioRedoxin; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 DR Redox-active center; Electron transport; Complete proteome.
 FT DISULFID 28 31 REDOX-ACTIVE (BY SIMILARITY).
 SQ SEQUENCE 102 AA; 11147 MW; C171B646D393428C CRC64;

Query Match 1.0%; Score 7; DB 1; Length 102;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 ASPRAE 228
 Db 57 ASPRAE 63

RESULT 9

ID R32A SCHPO STANDARD; PRT; 127 AA.
 AC P79015;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L32-A.
 GN RPL32A OR SPAC3H5.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares K., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey P., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomycetes pombe."
 RL Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 4-127 FROM N.A.
 RA Kawamukai M.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- MISCELLANEOUS: THERE ARE TWO GENES FOR L32 IN S.POMBE.
 CC -!- SIMILARITY: BELONGS TO THE L32E FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC -----
 DR EMBL; Z99296; CAB16594.1; --
 DR EMBL; AB000914; BAA19212.1; --
 DR InterPro; IPR001515; Ribosomal L32E.
 DR Pfam; PF01655; Ribosomal L32e; 1.
 DR ProDom; PD003823; Ribosomal L32E; 1.
 DR PROSITE; PS00580; RIBOSOMAL_L32E; 1.
 KW Ribosomal protein; Multigene family.
 SQ SEQUENCE 127 AA; 14458 MW; 4BBB3CEB2E359E05 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 127;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 569 NNKTRY 575
 Db 56 NNKTRY 62

RESULT 10

ID R32B SCHPO STANDARD; PRT; 127 AA.
 AC O42935;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L32-B.
 GN RPL32B OR SPAC16C.11.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares K., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey P., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomycetes pombe."
 RL Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 4-127 FROM N.A.
 RA Kawamukai M.;

RL Nature 415:871-880(2002).
CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L32 IN S.POMBE.
CC -1- SIMILARITY: BELONGS TO THE L32E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AL021767; CAA16918.1; -
DR InterPro; IPR001515; Ribosomal_L32E.
DR Pfam; PF01655; Ribosomal_L32e; 1.
DR PRODOM; PD003823; Ribosomal_L32e; 1.
DR PROSITE; PS00580; RIBOSOMAL_L32E; FALSE_NEG.
KW Ribosomal protein, multigene family.
SQ SEQUENCE 127 AA; 14445 MW; 0149CA472053669B CRC64;

Query Match 1.0%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 NNKTRY 575
Db 56 NNKTRY 62

RESULT 11

VG08_BPMU STANDARD; PRT; 139 AA.

AC Q38480;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein gp8 (Protein E7).

GN 8 OR E7.

OS Bacteriophage Mu.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;

OC Mu-like viruses.

OX NCBI_TaxId=10677;

RN [1]

RP SEQUENCE FROM N.A.

RA Prieses H., Brauer B., Schmidt C., Kamp D.;

RT "Sequence of the left end of Mu.:"

RL (In) Symonds N., Tounssaint A., van de Putte P., Howe M.M. (eds.);

RL Phage Mu, pp.277-296, Cold Spring Harbor Laboratory Press,

RL New York (1987).

RL [2]

RP SEQUENCE FROM N.A.

RA Morgan G., Hatfull G., Hendrix R.;

RT "Genome of bacteriophage Mu and comparison with the Haemophilus

RT influenzae Mu-like prophage PluMu.:"

RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL, M64097; AAA32391.1; -

DR EMBL, AF083977; AAF01134.1; -

KW Hypothetical protein.

SQ SEQUENCE 139 AA; 16310 MW; 80ACAC9B18F10A93 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 139;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 IPASIRE 72
Db 129 IPASIRE 135

RESULT 12

R27B_ARATH STANDARD; PRT; 146 AA.

AC Q9LR33;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 60S ribosomal protein L27a-B.

GN RPL27AB OR ATG623290 OR F26F24.13 OR F26F24_23.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucotyled II; Brassicales; Brassicaceae; Arabidopsie.

OX NCBI_TaxId=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,

RA White O., Alonso J., Altai H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chen Q., Chen H., Cheuk R.F., Chin C.W.,

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,

RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,

RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,

RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

RA Miltischer J., Miranda M., Nguyen M., Nielsen W.C., Osborne B.I.,

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,

RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,

RA Utecherback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,

RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

RT thaliana."

RL Nature 408:816-820(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;

RT "Riken Arabidopsis full length cDNA clones (RAFLs) sequenced by the

RT SSP consortium (Salk/Stanford/PGEC)."

RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 3-141 FROM N.A.

RC STRAIN=cv. Columbia;

RA Hofte H.;

RT Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL, AC005292; AAF68998.1; -

DR EMBL, AF349525; AAK15572.1; -

DR EMBL, AF324716; AAG40067.1; -

DR EMBL, AF410280; AAK95266.1; -

DR EMBL, Z26208; CAA81190.1; -

DR InterPro; IPR001196; Ribosomal_L15.

DR Pfam; PF00256; L15; 1.

DR PROSITE; PS00475; RIBOSOMAL_L15; FALSE_NEG.

KW Ribosomal protein.

or send an email to license@isb-sib.ch).

EMBL; X91959; CRA63025.1; -.
EMBL; AC011663; AAG52338.1; -.
EMBL; Z17767; CAA79059.1; -.
InterPro; IPR001196; Ribosomal_L15.
Pfam; PF00256; L15; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; FALSE_NEG.
KW Ribosomal protein.
SQ SEQUENCE 146 AA; 16455 MW; B04CB9121A0EE94C CRC64;

Query Match 1.0%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 GKVGMYR 318
|||||
DB 54 GKVGMYR 60

RESULT 13
R27C_ARATH STANDARD; PRT; 146 AA.

AC P49637;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L27a-C.
GN RPL27AC OR ATG1070600 OR F24J13.17 OR F5A18.22.
CC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eumetazoa II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97336304; PubMed=9193080;
RA Cooke R., Raynal M., Laudie M., Delsenly M.;
RT "Identification of members of gene families in Arabidopsis thaliana
by contig construction from partial cDNA sequences: 106 genes
encoding 50 cytoplasmic ribosomal proteins.";
RL Plant J. 11:1127-1140(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=1130712;
RA Theologis A., Ecker J.R., Palm C.U., Pedersen N.A., Kaul S.,
White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
Dunn P., Eggu P., Feldayyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
Gill J.E., Goldsmith A.D., Haas B., Hansen N.P., Hughes B., Huizar L.,
Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,
Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziali A.,
Millscher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
Utterback T., Van Aken S., Vaysberg M., Vyotskaia V.S., Walker M.,
Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
RN [3]
RP SEQUENCE OF 1-87 FROM N.A.
RC STRAIN=cv. Columbia;
RX Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delsenly M.;
RA Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X01573; CAA25729.1; -.
EMBL; M19490; AAA35002.1; -.
EMBL; Z72625; CAA96808.1; -.

CC This SWISS-PROT

```

DR EMBL; K01162; AAA35003.1; -.
DR PIR; A02782; R6B129.
DR SGD; S0003071; RPL28.
DR InterPro; IPR001196; Ribosomal_L15.
DR Pfam; PF00256; L15; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein; Cycloheximide resistance.
FT INIT_MET 0 0
FT DOMAIN 6 12 NUCLEAR LOCALIZATION SIGNAL.
FT DOMAIN 23 29 NUCLEAR LOCALIZATION SIGNAL.
FT VARIANT 37 37 E -> Q (CONFERES RESISTANCE TO
CYCLOHEXIMIDE, AN INHIBITOR OF
POLYPEPTIDE ELONGATION).
SQ SEQUENCE 148 AA; 16591 MW; F3E1CD53C6292FEAA CRC64;

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 148;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKGVMRY 318
Db 53 GKGVMRY 59

RESULT 15
RL2A_ERYGR STANDARD; PRT; 149 AA.
AC P78987;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L27a (L29).
OS Erythrae graminis (subsp. hordel) (Grass mildew).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
OC Erysiphales; Erysiphaceae; Blumeria.
OX NCBI_TaxID=62688;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR3;
RA Justesen A.F., Christiansen S.K., Giese H.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11394; CAA72204.1; -.
DR InterPro; IPR001196; Ribosomal_L15.
DR Pfam; PF00256; L15; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein.
SQ SEQUENCE 149 AA; 16786 MW; 599E3BD3195171 CRC64;

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 149;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKGVMRY 318
Db 54 GKGVMRY 60

RESULT 16
ID Y024_NPYOP STANDARD; PRT; 166 AA.
AC O10259;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

```

DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 18.4 kDa protein (ORF44).
OS Oryza pseudotsugata multicausid polyhedrosis virus (OpNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohmann G.F.;
RT "The sequence of the Oryza pseudotsugata multicausid polyhedrosis virus genome."
RT polyhedrosis virus genome."
RL Virology 229:381-399(1997).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75930; AAC59043.1; -.
KW Hypothetical protein.
SQ SEQUENCE 166 AA; 18390 MW; 1D6C0F2F4F877DDF CRC64;

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 166;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AATVLA 20
Db 105 AATVLA 111

RESULT 17
ATPD_ECOLI STANDARD; PRT; 177 AA.
AC P00831;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase delta chain (EC 3.6.3.14).
GN ATP OR UNCH OR PAPA OR B3735 OR Z5233 OR ECS4677.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85121806; PubMed=6395859;
RA Walker J.E., Gay N.J., Saraste M., Eberle A.N.;
RT "DNA sequence around the Escherichia coli unc operon. Completion of
RT the sequence of a 17 kilobase segment containing asna, oric, unc,
RT glms and phos."
RL Biochem. J. 224:799-815(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=82059437; PubMed=6272190;
RA Gay N.J., Walker J.E.;
RT "The atp operon: nucleotide sequence of the promoter and the genes
RT for the membrane proteins, and the delta subunit of Escherichia coli
RT ATP-synthase."
RL Nucleic Acids Res. 9:3919-3926(1981).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=82147764; PubMed=6278247;
RA Nielsen J., Hansen F.G., Hoppe J., Friedl P., von Meyenburg K.;
RT "The nucleotide sequence of the atp genes coding for the F0 subunits
RT a, b, c and the F1 subunit delta of the membrane bound ATP synthase

```

of *Escherichia coli*.";
RL Mol. Gen. Genet. 184:33-39(1981).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=82068433; PubMed=6458296;
RA Mabuchi K., Kanazawa H., Kayano T., Futai M.;
RT "Nucleotide sequence of the gene coding for the delta subunit of
RT proton translocating ATPase of *Escherichia coli*.";
RL Biochem. Biophys. Res. Commun. 102:172-179(1981).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the *Escherichia coli*
RT genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
RL Nature 409:529-533(2001).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
RA Kuwara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
RN [8]
RP STRUCTURE BY NMR OF 1-135.
RX MEDLINE=97307253; PubMed=9164460;
RA Wilkens S., Dunn S.D., Chandler J., Dahlquist F.W., Capaldi R.A.;
RT "Solution structure of the N-terminal domain of the delta subunit of
RT the E. coli ATP synthase.";
RL Nat. Struct. Biol. 4:198-201(1997).
RN [9]
RP FUNCTION: THIS PROTEIN SEEMS TO BE PART OF THE STALK THAT LINKS
CC CF(0) TO CF(1). IT EITHER TRANSMITS CONFORMATIONAL CHANGES FROM
CC CF(0) INTO CF(1) OR IS IMPLICATED IN PROTON CONDUCTION.
CC -|- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -|- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -|- SIMILARITY: BELONGS TO THE ATPASE DELTA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01594; AAA24734.1; -
DR EMBL; M12212; AAA20044.1; -
DR EMBL; X01631; CAA25779.1; -
DR EMBL; V00266; CAA23524.1; -
DR EMBL; V00264; CAA23517.1; -
DR EMBL; M25464; AAA83872.1; -
DR EMBL; L10328; AAA62087.1; -

DR EMBL; AE000450; AAC76758.1; -
DR EMBL; AE005605; AAG58938.1; -
DR EMBL; AP002566; BAB38100.1; -
DR PIR; A01031; PWECD.
DR POB; IABV; 07-JUL-97.
DR EcoGene; EG10105; atpH.
DR InterPro; IPR000711; ATPsynth_OSCP.
DR Pfam; PF00213; OSCF; 1.
DR PRINTS; PR00125; ATPASEDELTA.
DR TIGRFAMs; TIGR01145; ATP synt delta; 1.
DR PROSITE; PS00389; ATPASE_DELTA; 1.
KW Hydrolase; ATP synthesis; CF(1); Hydrogen ion transport; 3D-structure;
KW Complete proteome.
FT CONFLICT 82 82 E -> D (IN REF. 4).
SQ SEQUENCE 177 AA; 19332 MW; 6AFD9552A79C81A5 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 177;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
Db 161 DGSVRGR 167
|||||
|||

RESULT 18
RGSH CHICK
ID RGS CHICK STANDARD; PRT; 210 AA.
AC O9PWA0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Regulator of G-protein signaling 17 (RGS17).
GN RGS17.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Dorsal root ganglion;
RX MEDLINE=99348269; PubMed=10419452;
RA Jordan J.D., Carey K.D., Stork P.J.S., Iyengar R.;
RT "Modulation of Rap activity by direct interaction of Galphao with Rap1
RT GTPase-activating protein.";
RL J. Biol. Chem. 274:21507-21510(1999).
RN [2]
CC -|- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC THEIR INACTIVE GDP-BOUND FORM.
CC -|- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -|- PTM: FATTY ACYLATED. HEAVILY PALMITOYLATED IN THE CYSTEINE STRING
CC MOTIF (BY SIMILARITY).
CC -|- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF151968; AAD45948.1; -
DR HSP; P49795; 1CMZ.
DR InterPro; IPR000342; Regl_Gprotein.
DR Pfam; PF00615; RGS; 1.
DR PRINTS; PR01301; RGS-PROTEIN.
DR PRODOM; PD001580; Reg_of_prg; 1.
DR SMART; SM00315; RGS; 1.
DR PROSITE; PS50132; RGS; 1.
KW Signal transduction inhibitor; Membrane; Lipoprotein; Palmitate.
FT DOMAIN 28 40 POLY-CYS.

```

FT DOMAIN 84 200 RGS.
SQ SEQUENCE 210 AA; 24326 MW; 6581AD5BADDE7C CRC64;

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 210;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 MKTPAGR 342
   |||||
   |||||
Db 89 MKTPAGR 95

RESULT 19
RGS_MOUSE STANDARD; PRT; 210 AA.
ID RGS_MOUSE
AC 090280;
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Regulator of G-protein signaling 17 (RGS17) (Regulator of Gz-selective
DE protein signaling 2).
GN RGS17 OR RGS22.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C;
RA Barker S.A., Ross E.M.;
RT "RGS22, a new member of the Gz-selective GAP family.";
RL "RGS22, a new member of the Gz-selective GAP family."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=olfactory bulb;
RA MEDLINE=2108560; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Fukunishi Y., Komuro H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kankawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Holtman M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Schenwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Bois A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RA Nature 409:685-690(2001).
CC -1- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DIVING THEM INTO
CC THEIR INACTIVE GDP-BOUND FORM.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -1- PTM: FATTY ACYLATION. HEAVILY PALMITOYLATED IN THE CYSTEINE-STRING
CC MOTIF (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF191555; AAF05758.1; -.

```

```

DR EMBL; AK018279; BAB31145.1; -.
DR HSSP; P49795; 1CMZ.
DR MGD; MGI:1927469; Rgs17.
DR InterPro; IPR000342; Regl_Gprotein.
DR Pfam; PF00615; RGS; 1.
DR PRINTS; PR01301; RGS-PROTEIN.
DR ProDom; PD001580; Reg_of_prg; 1.
DR SMART; SM00315; RGS; 1.
DR PROSITE; PS50132; RGS; 1.
KW Signal transduction inhibitor; Membrane; Lipoprotein; Palmitate.
FT DOMAIN 28 40 POLY-CYS.
FT CHAIN 24 211
SQ SEQUENCE 210 AA; 24345 MW; 451B868679E5B9B0 CRC64;

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 210;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 MKTPAGR 342
   |||||
   |||||
Db 89 MKTPAGR 95

RESULT 20
ROPB_RHILV STANDARD; PRT; 211 AA.
ID ROPB_RHILV
AC 052866;
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 15-DEC-1998 (Rel. 37; Last annotation update)
DE 22 kDa outer membrane protein precursor.
GN ROPB.
OS Rhizobium leguminosarum (biovar viciae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=248;
RA MEDLINE=96172747; PubMed=8589412;
RA Roest H.P., Mulders I.H.M., Wijffelman C.A., Lugtenberg B.J.J.;
RT "Isolation of ropb, a gene encoding a 22-kDa Rhizobium leguminosarum
RT outer membrane protein.";
RL Mol. Plant Microbe Interact. 8:576-583(1995).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80767; CAA56751.1; -.
DR InterPro; IPR000498; OMPA_tmcm.
DR Pfam; PF01389; OMPA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 211
SQ SEQUENCE 211 AA; 22537 MW; 01925B81FB458228 CRC64;

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 GSVGRGV 215
   |||||
   |||||
Db 115 GSVGRGV 121

RESULT 21

```

```
ET3_MOUSE
ID ET3_MOUSE STANDARD; PRT; 214 AA.
AC P48299;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endothelin-3 precursor (ET-3).
GN EDN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RX MEDLINE=95094302; PubMed=8001160;
RA Baynash A.G., Hosoda K., Gaid A., Richardson J.A., Emoto N.,
RA Hammer R.E., Yanagisawa M.;
RT "Interaction of endothelin-3 with endothelin-B receptor is essential
RT for development of epidermal melanocytes and enteric neurons.";
RL Cell 79:1277-1285 (1994).
CC -!- FUNCTION: ENDOTHELINS ARE ENDOTHELIUM-DERIVED VASOCONSTRICTOR
CC PEPTIDES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE ENDOTHELIN/SARAFOTOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32330; AAB60509.1; -.
CC HSP; P05305; 1EDN.
CC MGD; MGI:95285; Edn3.
CC InterPro; IPR003641; Endothelin.
CC InterPro; IPR001928; Endothelin.tox.
CC Pfam; PF00322; endothelin; 1.
CC PRINTS; PR00365; ENDOTHELIN.
CC ProDom; PD005286; Endothelin; 1.
CC SMART; SM00272; END; 2.
CC PROSITE; PS00270; ENDOTHELIN; 2.
CC Cleavage on pair of basic residues; Vasoconstrictor; Multigene family;
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT PEPTIDE 97 117 ENDOTHELIN-3.
FT DOMAIN 159 173 ENDOTHELIN LIKE.
FT DISULFID 97 111 BY SIMILARITY.
FT DISULFID 99 107 BY SIMILARITY.
SQ SEQUENCE 214 AA; 23322 MW; 8911260872D6A713 CRC64;
Query Match 1.0%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 348 ADDKACA 354
DB 164 ADDKACA 170
|||||
RESULT 22
OMPW_VIBCH STANDARD; PRT; 217 AA.
AC P17266; O9KLR0;
DT 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein W precursor.
GN OMPW OR VCA0867.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=90245674; PubMed=2336399;
RA Jalajakumari M.B., Manning P.A.;
RT "Nucleotide sequence of the gene, ompW, encoding a 22kDa immunogenic
RT outer membrane protein of Vibrio cholerae.";
RL Nucleic Acids Res. 18:2180-2180 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA McDaniel L.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA Emdonald T., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: TO E.COLI OMPW AND P.OLEOVORANS ALK.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51948; CAA36210.1; -.
CC EMBL; AE004414; AAF96765.1; -.
CC PIR; S09509; S09509.
CC TIGR; VCA0867; -.
KW Signal; Outer membrane; Antigen; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 217 OUTER MEMBRANE PROTEIN W.
FT CONFLICT 7 7 MISSING (IN REF. 1).
FT CONFLICT 15 16 SS -> LA (IN REF. 1).
FT CONFLICT 59 59 Q -> H (IN REF. 1).
FT CONFLICT 79 79 A -> R (IN REF. 1).
FT CONFLICT 122 122 F -> N (IN REF. 1).
FT CONFLICT 142 142 G -> S (IN REF. 1).
FT CONFLICT 147 148 AG -> NA (IN REF. 1).
FT CONFLICT 179 179 S -> Y (IN REF. 1).
FT CONFLICT 210 210 M -> I (IN REF. 1).
SQ SEQUENCE 217 AA; 23316 MW; BF376BB11EF2797B CRC64;
Query Match 1.0%; Score 7; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 VLAALSS 23
DB 10 VLAALSS 16
|||||
RESULT 23
SPSA_BACSU STANDARD; PRT; 256 AA.
ID SPSA_BACSU
AC P39621;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Spore coat polysaccharide biosynthesis protein spsa.
GN SPSA OR IPA-63D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
```


RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,
 RA Presecan E., Santicu M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.,
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees.",
 RL Mol. Microbiol. 10:371-384(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
 RA Borriss R., Bourrier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Eutlan K.D., Evington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Katamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Pirescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scofield F.,
 RA Setiguchi J., Sekoweka A., Serron S.J., Serron P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takanashi A., Tanaka T., Terpestra P., Tognoni K.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Wellzenegger T.,
 RA Winiers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.",
 RL Nature 390:249-256(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RX MEDLINE=99280694; PubMed=10350455;
 RA Charneck S.J., Davies G.J.,
 RT "Structure of the nucleotide-diphospho-sugar transferase, Spas from
 RT Bacillus subtilis, in native and nucleotide-complexed forms.",
 RL Biochemistry 38:6380-6385(1999).
 CC -!- FUNCTION: GLYCOSYLTRANSFERASE IMPLICATED IN THE SYNTHESIS OF THE
 CC SPORE COAT.
 CC -!- PATHWAY: Spore coat polysaccharide biosynthesis.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, X73124; CA51619.1; -;
 DR EMBL, 299123; CAB15817.1; -;
 DR PDB, 1OG8; 21-APR-00.
 DR PDB, 1OGQ; 04-MAY-00.
 DR PDB, 1OGS; 04-MAY-00.
 DR Subtilast; BG10609; spsa.
 DR InterPro; IPR001173; Glycosyltransferase_2.
 DR Pfam; PF00535; Glycosyltransferase; 3D-structure; Complete proteome.
 KW

FT DISULFID 155 243
 SQ SEQUENCE 256 AA; 30184 MW; C20EA9627F5D536B CRC64;
 Query Match
 Best Local Similarity 1.0%; Score 7; DB 1; Length 256;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 572 KTRVVAL 578
 |||||
 DB 74 KTRVVAL 80
 RESULT 24
 ID ISPE_AQUAE STANDARD; PRT; 268 AA.
 AC 067060;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 41, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK
 DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 GN ISPE OR AQ.915.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
 OC Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.",
 RL Nature 392:353-358(1998).
 CC -!- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY
 CC SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC D-erythritol.
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
 CC step.
 CC -!- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AE000713; AAC07027.1; -;
 DR InterPro; IPR001745; GMPkinase_ATP.
 DR InterPro; IPR004424; ISPE.
 DR Pfam; PF00288; GMP_kinases; 1.
 DR TIGRFAMs; TIGR00154; ispe; 1.
 KW Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
 KW Complete proteome.
 FT NP BIND 88
 SO SEQUENCE 268 AA; 29788 MW; C905F8B98E8E0845D CRC64;
 Query Match
 Best Local Similarity 1.0%; Score 7; DB 1; Length 268;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 586 VETRIISG 592
 |||||
 DB 152 VETRIISG 158

```
RESULT 25
TEHB HAEIN STANDARD; PRT; 286 AA.
AC P45134;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tellurite resistance protein tehB homolog.
GN TEHB OR H11275.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512 (1995).
CC -!- FUNCTION: INVOLVED IN POTASSIUM TELLURITE RESISTANCE (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32807; AAC2923.1; -.
DR TIGR; H11275; -.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR004537; TehB.
DR TIGRFAMs; TIGR00477; tehB; 1.
KW Antibiotic resistance; Tellurium resistance; Complete proteome.
SQ SEQUENCE 286 AA; 33019 MW; 69D7647E97CC74FA CRC64;

Query Match 1.0%; Score 7; DB 1; Length 286;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 NSLYLSL 510
Db 134 NSLYLSL 140
|||||

RESULT 26
NADC SYN3
ID NADC SYN3 STANDARD; PRT; 295 AA.
AC P74301;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable nicotinate-nucleotide pyrophosphorylase [carboxylating]
DE (EC 2.4.2.19) (Quinolinate phosphoribosyltransferase
DE [decarboxylating]) (QAPrase).
GN NADC OR SLR0936.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 1.0%; Score 7; DB 1; Length 286;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 NSLYLSL 510
Db 134 NSLYLSL 140
|||||

RESULT 27
FEPB ECOLI STANDARD; PRT; 318 AA.
AC P14609;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferrienterobactin-binding periplasmic protein precursor.
GN FEPB OR B0592.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90008779; PubMed=2529253;
RA Elkins M.F., Eazhart C.F.;
RT "Nucleotide sequence and regulation of the Escherichia coli gene for
RT ferrienterobactin transport protein FepB.";
RL J. Bacteriol. 171:5443-5451 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
```

```

RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12.";
RN  Science 277:1453-1474(1997).
[3]
RP  SEQUENCE FROM N.A.
RA  Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA  Federspiel N., Hyman R., Kalkan S., Komp C., Kuri O., Lew H., Lin D.,
RA  Nameth A., Oetner P., Roberts D., Schramm S., Davis R.W.;
RN  Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[4]
RX  MEDLINE=90230305; PubMed=2139473;
RA  Brickman T.J., Odenberger B.A., McIntosh M.A.;
RT  "Regulation of divergent transcription from the iron-responsive fepB-
RT  entc promoter-operator regions in Escherichia coli.";
RN  J. Mol. Biol. 212:669-682(1990).
[5]
RP  CHARACTERIZATION.
RX  MEDLINE=96004464; PubMed=7551033;
RA  Stephens D.L., Choe M.D., Earhart C.F.;
RT  "Escherichia coli periplasmic protein fepB binds ferriterobactin.";
RN  Microbiology 141:1647-1654(1995).
CC  -1- FUNCTION: BINDS FERRITEROBACTIN. PART OF THE BINDING-PROTEIN-
CC  DEPENDENT TRANSPORT SYSTEM FOR UPTAKE OF FERRITEROBACTIN.
CC  -1- SUBCELLULAR LOCATION: Periplasmic.
CC  -1- INDUCTION: CONTROLLED IN PART BY THE AMOUNT OF AVAILABLE IRON.
CC  -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC  PROTEIN FAMILY 8.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
CC  DR  EMBL; M29730; AAA83853.1; -
CC  DR  EMBL; AE000164; AAC73693.1; -
CC  DR  EMBL; U82598; AAB40791.1; -
CC  DR  EMBL; X53274; CAA37370.1; -
CC  DR  PIR; JY0045; JY0045.
CC  DR  Ecocore; EG10294; fepB.
CC  DR  InterPro; IPR002491; Peripla_BP_2.1.
CC  DR  Pfam; PF01497; Peripla_BP_2.1.
CC  DR  Transprot; Iron transport; Signal; Periplasmic; Complete proteome.
CC  FT  SIGNAL 1 26 POTENTIAL.
CC  FT  CHAIN 27 318 FERRITEROBACTIN-BINDING PERIPLASMIC
CC  FT  PROTEIN.
CC  SQ  SEQUENCE 318 AA; 34283 MW; 807676D03A508079 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 318;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 HSNALMT 635
DB 204 HSNALMT 210

```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Em70;
RX  MEDLINE=21145866; PubMed=11248100;
RA  May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT  "Complete genomic sequence of Pasteurella multocida Em70.";
RN  Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC  -1- CATALYTIC ACTIVITY: Anthranilate + phosphoribosylpyrophosphate =
CC  N-5'-phosphoribosyl-anthranilate + diphosphate.
CC  -1- PATHWAY: Tryptophan biosynthesis; second step.
CC  -1- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
CC  FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
CC  DR  EMBL; AE006093; AAK02665.1; -
CC  DR  InterPro; IPR000312; Glycose_transf_3.
CC  DR  Pfam; PF00591; Glycose_transf_3.1.
CC  DR  Pfam; PF02885; Glycose_transf_3N.1.
CC  DR  ProDom; PD001864; Glycose_transf_3.1.
CC  DR  ProDom; PD005916; Thymid_phosphils; 1.
CC  DR  TIGRFAMs; TIGR01245; trpD; 1.
CC  KM  Tryptophan biosynthesis; Transferase; Glycosyltransferase;
CC  Complete proteome.
CC  SQ  SEQUENCE 333 AA; 35880 MW; F42760CD105F6C5A CRC64;

Query Match 1.0%; Score 7; DB 1; Length 333;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 589 EISGAVT 595
DB 54 EISGAVT 60

RESULT 29
ID  OTCC_HAEIN STANDARD; PRT; 334 AA.
AC  P44770;
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Ornithine carbamoyltransferase, catabolic (EC 2.1.3.3) (OTCasee).
GN  ARCB OR HI0596.
OS  Haemophilus influenzae.
OC  Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC  Haemophilus.
OX  NCBI_TaxID=727;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Rd / KW20 / ATCC 51907;
RX  MEDLINE=95550630; PubMed=7542800;
RA  Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA  Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA  Scott J.D., Shiley R., Liu L.-I., Glodok A., Kelley J.M.,
RA  Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA  Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA  Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
RA  Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA  Venter J.C.;
RT  "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT  Rd.";
RN  Science 269:496-512(1995).
CC  -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-ornithine = phosphate
CC  + L-citrulline.
CC  -1- PATHWAY: Arginine degradation via arginine deiminase; second step.

```

```

CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32741; AAC22253.1; -.
CC HSSP: P08308; 1ORT.
CC TIGR: H10596; -.
CC InterPro: IPR002029; Asp/Orn Cotransf.
CC InterPro: IPR002292; Orn cotransf.
CC Pfam: PF00185; OTCace; 1.
CC Pfam: PF02729; OTCace_N; 1.
CC PRINTS: PR00100; AOTCASE.
CC TIGRFAMs: TIGR00658; orni_carb_tr; 1.
CC TRANSFAM: TIGR00658; orni_carb_tr; 1.
KW Transferase: Arginine metabolism; Complete proteome.
SQ SEQUENCE 334 AA; 37647 MW; B44B9ED2C1B857F5 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 YQVTPPEL 643
DB 254 YQVTPPEL 260

RESULT 30
YAAA_SCHPO
ID YAAA_SCHPO STANDARD; PRT; 344 AA.
AC Q09801;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C2267.10 in chromosome I.
GN SPAC2267.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]_TaxID=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harzid P., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc K., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lenhard R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

```

```

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe I., McCombie W.R., Paulsen I., Potashkin J.,
RA Spakovski G.V., Uesery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: TO YEAST FIP1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z54328; CAA91134.1; -.
CC Hypothetical protein.
KW SEQUENCE 344 AA; 37290 MW; 571D391D0F88F27D CRC64;

Query Match 1.0%; Score 7; DB 1; Length 344;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 ITSSAGM 661
DB 272 ITSSAGM 278

RESULT 31
PHOE_CITFR
ID PHOE_CITFR STANDARD; PRT; 351 AA.
AC Q01605;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=546;
RN [1]_TaxID=546;
RP SEQUENCE FROM N.A.
RX MEDLINE=93146376; PubMed=1337052;
RA Spierings G., Ockhuijsen C., Hofstra H., Tommassen J.;
RT "Characterization of the Citrobacter freundii phoE gene and
RT development of C. freundii-specific oligonucleotides.";
RL FEMS Microbiol. Lett. 78:199-204(1992).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X68021; CAA48162.1; -.
CC PIR: S25520; S25520.
CC HSSP: P02932; IPHO.
CC InterPro: IPR003229; OMP_2.
CC Pfam: PF00267; Gram-ve porins; 1.
CC PRINTS: PR00182; ECOLNEIPORIN.
CC PROSITE: PS00576; GRAM_NEG_PORIN; 1.

```

KM Outer membrane; Transmembrane; Porin; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 351
 SQ SEQUENCE 351 AA; 38807 MW; 4E9C14A07031B5A CRC64;

Query Match 1.0%; Score 7; DB 1; Length 351;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 328 AFAGSKL 334
 |||||
 DB 98 AFAGSKL 104

RESULT 32
 MURB TREPA STANDARD; PRT; 354 AA.
 AC 083128;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE UDP-N-acetylglucosaminylglucosamine reductase (EC 1.1.1.158) (UDP-N-acetylglucosamine dehydrogenase).
 GN MURB OR TP0090.

OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 NC NCB1_Taxid=160;

[1]
 RP SEQUENCE FROM N.A.

RC MEDLINE=98332770; PubMed=9665876;
 RA Frazer C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artich P., Bowman C., Cotton M.D., Fujii C., Galand S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete.";
 RL Science 281:375-388(1998).

CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + NADP(+) = UDP-N-
 CC acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- PATHWAY: peptidoglycan biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: BELONGS TO THE MURB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: AE001193; AAC65086.1; -.
 DR TIGR: TP0090;
 DR InterPro: IPR003170; MurB.
 DR InterPro: IPR001575; Oxid_FAD_bind.
 DR Pfam: PF01565; FAD binding_4; 1.
 DR Pfam: PF02873; MurB_C; 1.
 KM Peptidoglycan synthetase; Cell wall; Cell division; Oxidoreductase;
 KM NADP; Flavoprotein; FAD; Complete proteome.
 SQ SEQUENCE 354 AA; 38595 MW; 809E5B606F5BFF2 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 354;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 LSLIGGG 467
 |||||

DB 56 LSLIGGG 62

RESULT 33
 V363 ASF7 STANDARD; PRT; 363 AA.
 AC P23164;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE D'363 protein.
 GN DP363R.
 OS African swine fever virus (strain BA71V) (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
 NC NCB1_Taxid=10498;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=90219205; PubMed=2325203;
 RA Gonzalez A., Calvo V., Almazan F., Almeral J.M., Ramirez J.C.,
 RA de la Vega I., Blasco R., Vinueza E.;
 RT "Multigene families in African swine fever virus: family 360.";
 RL J. Virol. 64:2073-2081(1990).

RP COMPLETE GENOME.
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 RA Rodriguez J.F., Vinueza E.;
 RT "Analysis of the complete nucleotide sequence of African swine fever
 RT virus.";
 RL Virology 208:249-278(1995).

CC -1- FUNCTION: NOT KNOWN.
 CC -1- SIMILARITY: BELONGS TO THE MULTIGENE FAMILY 360.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M57545; AAA42678.1; -.
 DR EMBL: U18466; AAA65382.1; -.
 DR PIR: F43680; F43680.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR002595; ASFV_360.
 DR Pfam: PF00023; ank; 1.
 DR Pfam: PF01671; ASFV_360; 1.
 DR PRODOM: PD003462; ASFV_360; 1.
 DR SMART; SM00248; ANK; 1.
 KM Multigene family.
 FT CONFLICT 241 M -> T (IN REF. 1).
 SQ SEQUENCE 363 AA; 42353 MW; 4841A8165447FE3 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 363;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 NLFAPDR 149
 |||||
 DB 349 NLFAPDR 355

RESULT 34
 IPAC SHIDY STANDARD; PRT; 382 AA.
 ID IPAC_SHIDY
 AC Q03946;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 42 kDa membrane antigen precursor.
 GN IPAC.
 OS Shigella dysenteriae.
 OG Plasmid invasion.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Shigella.
 OX NCBI_TaxID=622;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CG097;
 RX MEDLINE=921114800; PubMed=1766387;
 RA Yao R., Palchaudhuri S.;
 RT "Nucleotide sequence of the ipaBCD structural genes of Shigella
 RT dysenteriae.";
 RL Mol. Microbiol. 5:2217-2221(1991).
 CC -!- FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
 CC EPITHELIAL CELLS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X60777; CAA43191.1; -.
 DR PIR: S15578; S15578.
 KW Antigen; Plasmid; Virulence; Membrane; Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 382 42 KDA MEMBRANE ANTIGEN.
 SQ SEQUENCE 382 AA; 41098 MW; A22E1D6399EC90BF CRC64;
 Query Match 1.0%; Score 7; DB 1; Length 382;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 18 LAALSSS 24
 Db 170 LAALSSS 176
 RESULT 35
 IPAC_SHIFL STANDARD; PRT; 382 AA.
 AC P18012;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 42 kDa membrane antigen precursor.
 GN IPAC.
 OS Shigella flexneri.
 OG Plasmid 210 kb invasion pWR100, and plasmid 230 kb pMYSH6000.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Shigella.
 OX NCBI_TaxID=623;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=M90T / Serotype 5;
 RX MEDLINE=89057927; PubMed=3057506;
 RA Venkatesan M.M., Buysse J.M., Kopecko D.J.;
 RT "Characterization of invasion plasmid antigen genes (ipaBCD) from
 RT Shigella flexneri.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9317-9321(1988).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=M90T / Serotype 5;
 RX MEDLINE=89200844; PubMed=3071655;
 RA Baudry B., Kaczorek M., Sansonetti P.J.;
 RT "Nucleotide sequence of the invasion plasmid antigen B and C genes
 RT (ipaB and ipaC) of Shigella flexneri.";
 RL Microb. Pathog. 4:345-357(1988).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Serotype 2A; PLASMID=230 kb pMYSH6000;
 RX MEDLINE=90014179; PubMed=2552264;
 RA Sasakawa C., Adler B., Tobe T., Okada N., Nagai S., Komatsu K.,

RA Yoshikawa M.;
 RT "Functional organization and nucleotide sequence of virulence
 RT Region-2 on the large virulence plasmid in Shigella flexneri 2a.";
 RL Mol. Microbiol. 3:1191-1201(1989).
 RN (4)
 RP PARTIAL SEQUENCE OF 20-64 AND 318-335.
 RX MEDLINE=89307550; PubMed=2663721;
 RA Sankaran K., Ramachandran V., Subrahmanyam Y.V.B.K., Rajarathnam S.,
 RA Kango S., Roy R.K.;
 RT "Congo red-mediated regulation of levels of Shigella flexneri 2a
 RT membrane proteins.";
 RL Infect. Immun. 57:2364-2371(1989).
 CC -!- FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
 CC EPITHELIAL CELLS.
 CC -----
 CC -!- MISCELLANEOUS: SYNTHESIS OF THIS IMMUNOGEN IS RERESSED AT 30
 CC DEGREES CELSIUS AND RESTORED AT 37 DEGREES CELSIUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04117; AAA26523.1; -.
 DR EMBL: M34849; AAA98425.1; ALT INIT.
 DR EMBL: X15319; CAA33382.1; -.
 DR PIR: A60112; A60112.
 KW Antigen; Plasmid; Virulence; Membrane; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 382 42 KDA MEMBRANE ANTIGEN.
 FT CONFLICT 30 30 T -> I (IN REF. 3).
 FT CONFLICT 372 372 A -> T (IN REF. 3).
 SQ SEQUENCE 382 AA; 41038 MW; ED6481B161E58EB1 CRC64;
 Query Match 1.0%; Score 7; DB 1; Length 382;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 18 LAALSSS 24
 Db 170 LAALSSS 176
 RESULT 36
 BN51_HUMAN STANDARD; PRT; 395 AA.
 ID BN51_HUMAN
 AC P05423;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE BN51 protein.
 GN BN51T OR BN51.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88065472; PubMed=3693386;
 RA Ittmann M., Greco A., Basilico C.;
 RT "Isolation of the human gene that complements a temperature-sensitive
 RT cell cycle mutation in BHK cells.";
 RL Mol. Cell. Biol. 7:3386-3393(1987).
 RN (2)
 RP SEQUENCE OF 1-12 FROM N.A.
 RA Ittmann M.;
 RT "Promoter structure and cell cycle control of the BN51 cell cycle
 RT gene, which encodes a subunit of RNA polymerase III.";
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: COMPLEMENTS A TEMPERATURE-SENSITIVE CELL CYCLE MUTATION
 CC IN BHK CELLS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M17754; AAA51838.1; -
DR EMBL; L15301; AAA72377.1; -
DR PIR; A43700; A43700.
DR Genew; HGNC:1080; BMS1T.
DR MIM; 187280; -
FT DOMAIN 62 98 HYDROPHILIC.
FT DOMAIN 138 149 HYDROPHILIC.
FT DOMAIN 215 230 HYDROPHILIC.
FT DOMAIN 289 306 HYDROPHILIC.
FT SIMILAR 289 395 TO THE C-TERMINUS OF YEAST RPC4.
SQ SEQUENCE 395 AA; 43978 MW; 897DEC72571844D CRC64;

Query Match 1.0%; Score 7; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97 ARKTPGL 103
DB 236 ARKTPGL 242

RESULT 37
VGLX HSVB STANDARD; PRT; 411 AA.
ID P28967;
AC 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein G precursor.
GN 70.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; deDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_Taxid=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316(1992).
CC -1- SIMILARITY: TO PRV GLYCOPROTEIN X.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M86664; AAB02505.1; -
DR PIR; G36802; VGBEG2.
DR InterPro; IPR003363; Herpes_gg.
DR Pfam; PF02400; Herpes_gg; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 17
FT CHAIN 18 411 GLYCOPROTEIN G.
FT TRANSMEM 363 379 POTENTIAL.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 411 AA; 45269 MW; CC6349C38D0E12E2 CRC64;
Query Match 1.0%; Score 7; DB 1; Length 411;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 TVLAALS 22
DB 3 TVLAALS 9

RESULT 38
VGLX HSVB STANDARD; PRT; 411 AA.
ID P32514;
AC 01-OCT-1993 (Rel. 27, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycoprotein G precursor.
GN US3.
OS Equine herpesvirus type 1 (strain Kentucky A) (EHV-1).
OC Viruses; deDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_Taxid=10329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92263758; PubMed=1316673;
RA Colle C.F., Flowers C.C., O'Callaghan D.J.;
RT "Open reading frames encoding a protein kinase, homolog of
RT glycoprotein gX of pseudorabies virus, and a novel glycoprotein map
RT within the unique short segment of equine herpesvirus type 1.";
RL Virology 188:545-557(1992).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Kinkou M., Fukushi H., Matsumura T., Kim S.K., O'Callaghan D.J.;
RL Submitted (SEP-2001) to the EMBL/Genbank/DBD databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M87497; AAA46071.2; -
DR PIR; B42538; VGBEKA.
DR InterPro; IPR003363; Herpes_gg.
DR Pfam; PF02400; Herpes_gg; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 19
FT CHAIN 20 411
FT TRANSMEM 364 384 POTENTIAL.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 411 AA; 45327 MW; CESDB8348CACC1144 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 411;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 TVLAALS 22
DB 3 TVLAALS 9

RESULT 39
GLUP BRUB STANDARD; PRT; 412 AA.
ID GLUP BRUB
AC Q44623;

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Glucose/galactose transporter.
 GN GLUP.
 OS Brucella abortus.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=235;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2308;
 RX MEDLINE=97311994; PubMed=9168605;
 RA Essenberg R.C., Candler C., Nida S.K.;
 RT "Brucella abortus strain 2308 putative glucose and galactose
 transporter gene: cloning and characterization.";
 RL Microbiology 143:1549-1555(1997).
 CC -!- FUNCTION: INTAKE OF GLUCOSE AND GALACTOSE (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -!- SIMILARITY: BELONGS TO THE FHS (TC 2.1.7) FAMILY OF TRANSPORTERS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U43785; AAB58958.1; -;
 DR TIGRFAMs; TIGR01272; Glup; 1.
 KW Sugar transport; Transmembrane; Inner membrane.
 FT TRANSMEM 21 41 POTENTIAL.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 113 133 POTENTIAL.
 FT TRANSMEM 158 178 POTENTIAL.
 FT TRANSMEM 192 212 POTENTIAL.
 FT TRANSMEM 239 259 POTENTIAL.
 FT TRANSMEM 310 330 POTENTIAL.
 FT TRANSMEM 331 351 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 388 408 POTENTIAL.
 SQ SEQUENCE 412 AA; 43859 MW; 716D3AAB9E197E5F CRC64;
 Query Match 1.0%; Score 7; DB 1; Length 412;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 364 KALAFDA 370
 DB 304 KALAFDA 310
 RESULT 40
 Y118 TREPA
 ID Y118 TREPA STANDARD; PRT; 421 AA.
 AC O83155;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein TP0118.
 GN TP0118.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 spirochete.";
 RL Science 281:375-388(1998).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE001197; AAC65114.1; -;
 DR TIGR; TP0118; -;
 KW Hypothetical protein; Coiled coil; Complete proteome.
 FT DOMAIN 126 182 COILED COIL (POTENTIAL).
 FT DOMAIN 328 397 COILED COIL (POTENTIAL).
 SQ SEQUENCE 421 AA; 46702 MW; 8C7C72BD0C515B6A CRC64;
 Query Match 1.0%; Score 7; DB 1; Length 421;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 576 AALGKRV 582
 DB 82 AALGKRV 88
 RESULT 41
 FOLC HAEIN
 ID FOLC HAEIN STANDARD; PRT; 437 AA.
 AC P43775;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Folylpolyglutamate synthase (EC 6.3.2.17) (Folylpoly-gamma-glutamate
 synthetase) (FPGS).
 DE FOLC OR H1261.
 GN Haemophilus influenzae.
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
 RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Uddelman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -!- FUNCTION: CONVERSION OF FOLATES TO POLYGLUTAMATE DERIVATIVES
 CC (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + {tetrahydrofolyl-[Glu]}(N) + L-glutamate
 CC = ADP + phosphate + {tetrahydrofolyl-[Glu]}(N+1).
 CC -!- PATHWAY: BACTERIA REQUIRE FOLATE FOR THE BIOSYNTHESIS OF GLYCINE,
 CC METHIONINE, FORMYL-MET-TRNA, THYMIDYLATES, PURINES, AND
 CC PANTOTHENATE.
 CC -!- SIMILARITY: BELONGS TO THE FOLYLPOLYGLUTAMATE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32806; AAC22914.1; -.
CC DR HSSP; P15925; 1FGS.
CC DR TIGR; H11261; -.
CC DR InterPro; IPR001645; Fpolylgl_synthase.
CC DR InterPro; IPR000713; Mur_ligase.
CC DR Pfam; PF01225; Mur_ligase; 1.
CC DR PROSITE; PS01011; FOLYLPOLYGLU_SYNT_1; 1.
CC DR PROSITE; PS01012; FOLYLPOLYGLU_SYNT_2; 1.
CC KW Ligase; One-carbon metabolism; ATP-binding; Complete proteome.
CC FT NP BIND 55 ATP (BY SIMILARITY).
CC SQ SEQUENCE 437 AA; 48826 MW; 04A493395E76CD43 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 437;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLA 20
Db 256 AATVLA 262

RESULT 42
UVSH_EMENT STANDARD; PRT; 443 AA.
AC Q02398; Q00178;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Postreplication repair protein uvsh/nuva.
GN UVSH OR NUVA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxId=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGSC 4;
RX MEDLINE=95379762; PubMed=7651340;
RA Yoon J.H., Lee B.J., Kang H.S.;
RT "The Aspergillus uvsh gene encodes a product homologous to yeast Rad18
RT and Neurospora Uvs-2."
RT Mol. Genet. 248:174-181(1995).
RL [2]
RN SEQUENCE OF 1-356 FROM N.A.
RP STRAIN=L20;
RC MEDLINE=97022065; PubMed=8868425;
RA Iwawake L.A., Cotton C.M., Jones G.W., Tomsett A.B., Strike P.;
RT "nuva", an Aspergillus nidulans gene involved in DNA repair and
RT recombination, is a homologue of Saccharomyces cerevisiae Rad18 and
RT Neurospora crassa uvs-2."
RT Microbiology 142:505-515(1996).
RL
CC -!- FUNCTION: Involved in postreplication repair of UV-damaged DNA.
CC Postreplication repair functions in gap-filling of a daughter
CC strand on replication of damaged DNA. Has ssDNA binding activity
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; S79392; AAB35098.1; -.
DR EMBL; Z49875; CAA90033.1; -.
DR HSSP; P15919; 1RMD.
DR InterPro; IPR004580; Rad18.
DR InterPro; IPR003034; SAP.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00513; SAP; 1.
DR TIGRfam; TIGR00599; rad18; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
DR Zinc-finger; DNA-binding; DNA damage; DNA repair; Nuclear protein.
FT ZN FING 30 68
FT DOMAIN 236 270 SAP.
FT CONFLICT 209 210 FG -> YR (IN RRF, 2).
SQ SEQUENCE 443 AA; 48906 MW; DD3327065D3511B2 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 443;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 LIANARK 430
Db 334 LIANARK 340

RESULT 43
SR54_AQUAE STANDARD; PRT; 454 AA.
ID SR54_AQUAE
AC O67615;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN PFH OR AQ_1720.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxId=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed W.A., Keller M., Anjay M., Huber R.,
RA Felman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RT Nature 392:353-358(1998).
RL
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN PFH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE. THE G-DOMAIN
CC BINDS GTP. THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000753; AAC07579.1; -.
CC HSSP; O07347; 1FRH.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR000897; SRP54.

```

```

DR InterPro; IPR004125; SRP54_SBP.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54_N; 1.
DR Pfam; PF02978; SRP_SBP; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; ARA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 289 G-DOMAIN.
FT DOMAIN 290 454 M-DOMAIN.
FT NP_BIND 102 109 GTP (BY SIMILARITY).
FT NP_BIND 184 188 GTP (BY SIMILARITY).
FT NP_BIND 242 245 GTP (BY SIMILARITY).
SQ SEQUENCE 454 AA; 50911 MW; 9F13FA8C883A0558 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 454;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 ARGGLAL 413
DB 248 ARGGLAL 254

RESULT 44
ID EF1A_DICDI STANDARD; PRT; 456 AA.
AC P18624;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor 1-alpha (EF-1-alpha) (50 kDa actin-binding protein)
DE (ABP-50).
DN EFPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=AX3;
RX MEDLINE=91015340; PubMed=2215665;
RA Yang F., Demma M., Warren V., Dharmawardane S., Condeelis J.;
RT "Identification of an actin-binding protein from Dictyostelium as
RT elongation factor 1a.";
RL Nature 347:494-496(1990).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS. IT IS ALSO AN ABUNDANT ACTIN FILAMENT BUNDLING
CC PROTEIN.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X55972; CAA39442.1; -.
DR EMBL; X55973; CAA39443.1; -.
DR PIR; S11665; S11665.
DR HSP; P07157; 1A1P.
DR DictyDB; DD01027; efaA.
DR InterPro; IPR004539; EF1_alpha.
DR InterPro; IPR004160; EFTU_Cterm.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.

```

```

DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03143; GTP_EFTU_D3; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR TIGRFAMs; TIGR00483; EF-1_alpha; 1.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding; Actin-binding.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 94 98 GTP (BY SIMILARITY).
FT NP_BIND 156 159 GTP (BY SIMILARITY).
SQ SEQUENCE 456 AA; 50066 MW; BCFE9E1B70C3D8A8 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 QIHAGYS 605
DB 343 QIHAGYS 349

RESULT 45
TBB1_PORPU STANDARD; PRT; 457 AA.
AC P50259;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tubulin beta-1 chain.
DN TUBB1.
OS Porphyra purpurea.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Mackay R.M., Gallant J.W.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC -!- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z67991; CAA91939.1; -.
DR InterPro; IPR002453; Beta tubulin.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_FtsZ.
DR Pfam; PF00091; tubulin; 1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
DR PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW Microtubules; GTP-binding; Multigene family.
FT NP_BIND 140 146 GTP (POTENTIAL).
SQ SEQUENCE 457 AA; 50047 MW; 0F2E5BFA7E077DAA CRC64;

Query Match 1.0%; Score 7; DB 1; Length 457;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 MSGITSS 658
DB 233 MSGITSS 239

```

Search completed: December 18, 2002, 06:57:25
Job time : 30.2194 secs

270

THIS PAGE BLANK (USPTO)

977 5 0.7 13 4 US-09-196-293-6 Sequence 6, Appli
 978 5 0.7 13 4 US-08-602-999A-88 Sequence 88, Appl
 979 5 0.7 13 4 US-08-602-999A-89 Sequence 89, Appl
 980 5 0.7 13 4 US-08-895-590-17 Sequence 17, Appl
 981 5 0.7 13 4 US-08-209-603B-6 Sequence 6, Appli
 982 5 0.7 13 4 US-08-278-865-88 Sequence 88, Appl
 983 5 0.7 13 4 US-08-278-865-89 Sequence 89, Appl
 984 5 0.7 13 4 US-08-278-865-90 Sequence 90, Appl
 985 5 0.7 13 4 US-09-500-124-88 Sequence 88, Appl
 986 5 0.7 13 4 US-09-500-124-89 Sequence 89, Appl
 987 5 0.7 13 5 PCT-US93-12246-4 Sequence 4, Appli
 988 5 0.7 13 5 PCT-US95-04121-10 Sequence 10, Appl
 989 5 0.7 13 5 PCT-US96-11495-2 Sequence 2, Appli
 990 5 0.7 13 5 PCT-US96-11495-22 Sequence 22, Appl
 991 5 0.7 14 1 US-08-036-555B-126 Sequence 126, App
 992 5 0.7 14 1 US-08-469-569-126 Sequence 126, App
 993 5 0.7 14 1 US-08-249-322A-126 Sequence 126, App
 994 5 0.7 14 1 US-08-642-255-103 Sequence 103, App
 995 5 0.7 14 1 US-08-397-633A-5 Sequence 5, Appli
 996 5 0.7 14 1 US-08-397-633A-10 Sequence 10, Appl
 997 5 0.7 14 1 US-08-469-526A-126 Sequence 126, App
 998 5 0.7 14 2 US-08-734-591A-126 Sequence 126, App
 999 5 0.7 14 2 US-08-637-759B-250 Sequence 250, App
 1000 5 0.7 14 2 US-08-469-660-126 Sequence 126, App

ALIGNMENTS

RESULT 1
 US-08-656-177A-3
 ; Sequence 3, Application US/08656177A
 ; Patent No. 5882851
 ; GENERAL INFORMATION:
 ; APPLICANT: Koch, Birgit M.
 ; APPLICANT: Sibbesen, Ole
 ; APPLICANT: Halkier, Barbara Ann
 ; APPLICANT: Moller, Birger L.
 ; TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5882851artis Corporation
 ; STREET: 3054 Cornwallis Road
 ; CITY: Research Triangle Park
 ; STATE: NC
 ; COUNTRY: USA
 ; ZIP: 27709
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/656,177A
 ; FILING DATE: 08-AUG-1996
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; REFERENCE/DOCKET NUMBER: S-19783/A/PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 17 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; US-08-656-177A-3

Query Match 1.0%; Score 7; DB 2; Length 17;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 14 AATVLAA 20
 Db 10 AATVLAA 16
 |||||

RESULT 2
 US-09-256-797-3
 ; Sequence 3, Application US/09256797
 ; Patent No. 6133417
 ; GENERAL INFORMATION:
 ; APPLICANT: Koch, Birgit M.
 ; APPLICANT: Sibbesen, Ole
 ; APPLICANT: Halkier, Barbara Ann
 ; APPLICANT: Moller, Birger L.
 ; TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6133417artis Corporation
 ; STREET: 3054 Cornwallis Road
 ; CITY: Research Triangle Park
 ; STATE: NC
 ; COUNTRY: USA
 ; ZIP: 27709
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/256,797
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/656,177
 ; FILING DATE: 08-AUG-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; REFERENCE/DOCKET NUMBER: S-19783/A/PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 17 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; US-09-256-797-3

Query Match 1.0%; Score 7; DB 4; Length 17;
 Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 14 AATVLAA 20
 Db 10 AATVLAA 16
 |||||

RESULT 3
 US-08-817-811-97
 ; Sequence 97, Application US/08817811
 ; Patent No. 6174528
 ; GENERAL INFORMATION:
 ; APPLICANT: Cooper, Juan A.

APPLICANT: Relf, Wendy A.
APPLICANT: Good, Michael F.
APPLICANT: Saul, Allan J.
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
TITLE OF INVENTION: COMPRISING SAME
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-97

Query Match 1.0%; Score 7; DB 4; Length 75;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 407 ARGJAL 413
Db 23 ARGJAL 29

RESULT 4
US-08-817-811-21
Sequence 21, Application US/08817811
Patent No. 6174528
GENERAL INFORMATION:
APPLICANT: Cooper, Juan A.
APPLICANT: Relf, Wendy A.
APPLICANT: Good, Michael F.
APPLICANT: Saul, Allan J.
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
TITLE OF INVENTION: COMPRISING SAME
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-21

Query Match 1.0%; Score 7; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 407 ARGJAL 413
Db 23 ARGJAL 29

RESULT 5
US-08-817-811-20
Sequence 20, Application US/08817811
Patent No. 6174528
GENERAL INFORMATION:
APPLICANT: Cooper, Juan A.
APPLICANT: Relf, Wendy A.
APPLICANT: Good, Michael F.
APPLICANT: Saul, Allan J.
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
TITLE OF INVENTION: COMPRISING SAME
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-20

Query Match 1.0%; Score 7; DB 4; Length 78;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 ARGGLAL 413
Db 27 ARGGLAL 33

RESULT 6

US-08-817-811-96
; Sequence 96, Application US/08817811
; Patent No. 6174528
; GENERAL INFORMATION:
; APPLICANT: Cooper, Juan A.
; APPLICANT: Relf, Wendy A.
; APPLICANT: Good, Michael F.
; APPLICANT: Saul, Allan J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
; TITLE OF INVENTION: COMPRISING SAME
; NUMBER OF SEQUENCES: 97
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 14-APR-1997
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-96

Query Match 1.0%; Score 7; DB 4; Length 78;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 ARGGLAL 413
Db 27 ARGGLAL 33

RESULT 7

US-08-817-811-19
; Sequence 19, Application US/08817811
; Patent No. 6174528
; GENERAL INFORMATION:

APPLICANT: Cooper, Juan A.
APPLICANT: Relf, Wendy A.
APPLICANT: Good, Michael F.
APPLICANT: Saul, Allan J.
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
TITLE OF INVENTION: COMPRISING SAME
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 79 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-19

Query Match 1.0%; Score 7; DB 4; Length 79;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 ARGGLAL 413
Db 30 ARGGLAL 36

RESULT 8

US-08-817-811-95
; Sequence 95, Application US/08817811
; Patent No. 6174528
; GENERAL INFORMATION:
; APPLICANT: Cooper, Juan A.
; APPLICANT: Relf, Wendy A.
; APPLICANT: Good, Michael F.
; APPLICANT: Saul, Allan J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
; TITLE OF INVENTION: COMPRISING SAME
; NUMBER OF SEQUENCES: 97
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 79 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-95

Query Match 1.0%; Score 7; DB 4; Length 79;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 407 ARGGLAL 413
|||
Db 30 ARGGLAL 36

RESULT 9
US-08-392-828C-39
Sequence 39, Application US/08392828C
Patent No. 5795362
GENERAL INFORMATION:
APPLICANT: IWANAGA, SADAKI
APPLICANT: MUTA, TATSUSHI
APPLICANT: SEKI, NORIAKI
APPLICANT: ODA, TOSHIO
TITLE OF INVENTION: NOVEL POLYPEPTIDE AND DNA ENCODING
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,828C
FILING DATE: 28-FEB-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, PAULA A
REGISTRATION NUMBER: 32,503
REFERENCE/DOCKET NUMBER: FPN-033
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..127
OTHER INFORMATION: /note= "XLN A SEQUENCE (FIGURE 4)"
US-08-392-828C-39

Query Match 1.0%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 205 LNSDGSV 211
|||
Db 83 LNSDGSV 89

RESULT 10
US-09-330-945-39
Sequence 39, Application US/09330945
Patent No. 6077946
GENERAL INFORMATION:
APPLICANT: IWANAGA, SADAKI
APPLICANT: MUTA, TATSUSHI
APPLICANT: SEKI, NORIAKI
APPLICANT: ODA, TOSHIO
TITLE OF INVENTION: DNA ENCODING HORSESHOE CRAB
TITLE OF INVENTION: AMEBOCYTE LYSATE FACTOR G SUBUNIT A
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/119,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: FPN-032DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..127
OTHER INFORMATION: /note= "XLN A SEQUENCE (FIGURE 4)"
US-09-330-945-39

Query Match 1.0%; Score 7; DB 3; Length 127;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 LNSDGSV 211
Db 83 LNSDGSV 89

RESULT 11

5182262-8
; Patent No. 5182262
; APPLICANT: LETO, THOMAS
; TITLE OF INVENTION: CALMODULIN BINDING PEPTIDE DERIVATIVES
; OF NON-ERYTHROID ALPHA SPECTRIN
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,172
; FILING DATE: 02-MAR-1989
; SEQ ID NO: 8:
; LENGTH: 248
5182262-8

Query Match 1.0%; Score 7; DB 6; Length 248;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 HTVATEN 443
Db 139 HTVATEN 145

RESULT 12

US-07-941-651-2
; Sequence 2, Application US/07941651
; Patent No. 5332808
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5332808th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/941,651
; FILING DATE: 19920908

CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-941-651-2

Query Match 1.0%; Score 7; DB 1; Length 278;

Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 558 KDKNAAA 564
Db 262 KDKNAAA 268

RESULT 13

US-08-279-996-2
; Sequence 2, Application US/08279996
; Patent No. 5552140
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5552140th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/279,996
; FILING DATE: 25-JUL-1994

CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/941,651
; FILING DATE: 08-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102

INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-279-996-2

Query Match 1.0%; Score 7; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 558 KDKNAAA 564
Db 262 KDKNAAA 268

RESULT 14

US-08-746-797-2
; Sequence 2, Application US/08746797
; Patent No. 5759832
; GENERAL INFORMATION:
; APPLICANT: Gentry, Daniel
; APPLICANT: Lonsdale, John
; APPLICANT: Payne, David
; APPLICANT: Pearson, Stewart

;; TITLE OF INVENTION: NOVEL FabH
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SmithKline Beecham Corporation
;; STREET: 709 Swedeland Road
;; CITY: King of Prussia
;; STATE: PA
;; COUNTRY: U.S.A.
;; ZIP: 19406-0939
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 1.5
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/746,797
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: No. 5759832 Yet Assigned
;; FILING DATE: 23-OCT-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gimmi, Edward R
;; REGISTRATION NUMBER: 38,891
;; REFERENCE/DOCKET NUMBER: P50573
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 610-270-4478
;; TELEFAX: 610-270-5090
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 324 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHEICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: N-terminal
;; ORIGINAL SOURCE:
US-08-746-797-2

Query Match 1.0%; Score 7; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 SLNSDGS 210
DB 178 SLNSDGS 184

RESULT 15
US-08-927-387-2
;; Sequence 2, Application US/08927387
;; Patent No. 5783432
;; GENERAL INFORMATION:
;; APPLICANT: Gentry, Daniel
;; APPLICANT: Lonsdale, John
;; APPLICANT: Payne, David
;; APPLICANT: Pearson, Stewart
;; TITLE OF INVENTION: NOVEL FabH
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SmithKline Beecham Corporation
;; STREET: 709 Swedeland Road
;; CITY: King of Prussia
;; STATE: PA
;; COUNTRY: U.S.A.
;; ZIP: 19406-0939
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; OPERATING SYSTEM: DOS

;; SOFTWARE: FastSeq Version 1.5
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/927,387
;; FILING DATE: 25-AUG-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/746,797
;; FILING DATE: 23-OCT-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gimmi, Edward R
;; REGISTRATION NUMBER: 38,891
;; REFERENCE/DOCKET NUMBER: P50573
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 610-270-4478
;; TELEFAX: 610-270-5090
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 324 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHEICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: N-terminal
;; ORIGINAL SOURCE:
US-08-927-387-2

Query Match 1.0%; Score 7; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 SLNSDGS 210
DB 178 SLNSDGS 184

RESULT 16
US-08-918-058-2
;; Sequence 2, Application US/08918058
;; Patent No. 5885572
;; GENERAL INFORMATION:
;; APPLICANT: Gentry, Daniel
;; APPLICANT: Lonsdale, John
;; APPLICANT: Payne, David
;; APPLICANT: Pearson, Stewart
;; TITLE OF INVENTION: NOVEL FabH
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SmithKline Beecham Corporation
;; STREET: 709 Swedeland Road
;; CITY: King of Prussia
;; STATE: PA
;; COUNTRY: U.S.A.
;; ZIP: 19406-0939
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 1.5
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/918,058
;; FILING DATE: 25-AUG-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/746,797
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gimmi, Edward R
;; REGISTRATION NUMBER: 38,891
;; REFERENCE/DOCKET NUMBER: P50573
;; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 324 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-918-058-2

Query Match 1.0%; Score 7; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 204 SLNSDGS 210
Db 178 SLNSDGS 184

RESULT 17

US-08-921-887-52
; Sequence 52, Application US/08921887
; Patent No. 6030771
; GENERAL INFORMATION:
; APPLICANT: KHUYAKOV, YURI B.
; APPLICANT: FIELDS, HOWARD A.
; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JONES & ASKEW, LLP
; STREET: 191 Peachtree Street, N.W., 37th Floor
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,887
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WARREN, WILLIAM L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 03063-0380
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis virus
; US-08-921-887-52

Query Match 1.0%; Score 7; DB 3; Length 352;

Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 KQKALAF 368
Db 60 KQKALAF 66

RESULT 18

US-08-338-530A-3
; Sequence 3, Application US/08338530A
; Patent No. 5922327
; GENERAL INFORMATION:
; APPLICANT: CRABB, Brendan S.
; APPLICANT: STUDDERT, Michael J.
; TITLE OF INVENTION: EQUINE HERPESVIRUS GLYCOPROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/338,530A
; FILING DATE: 25-JAN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/AU93/00253
; FILING DATE: 28-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40268/120/CSMB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-338-530A-3

Query Match 1.0%; Score 7; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 TVLAALS 22
Db 3 TVLAALS 9

RESULT 19

US-09-267-384-3
; Sequence 3, Application US/09267384
; Patent No. 6193983
; GENERAL INFORMATION:
; APPLICANT: CRABB, Brendan S.
; APPLICANT: STUDDERT, Michael J.
; TITLE OF INVENTION: EQUINE HERPESVIRUS GLYCOPROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/267,384
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/338,530
FILING DATE: 25-JAN-1995
APPLICATION NUMBER: WO PCT/AU93/00253
FILING DATE: 28-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40268/120/CSMB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-267-384-3

Query Match 1.0%; Score 7; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TVLAALS 22
Db 3 TVLAALS 9

RESULT 20
US-09-457-046B-64
; Sequence 64, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Pacifitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 64
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-64

Query Match 1.0%; Score 7; DB 4; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 668 TFDMAA 674
Db 277 TFDMAA 283

RESULT 21
US-08-468-812-8
; Sequence 8, Application US/08468812

Patent No. 5935836
; GENERAL INFORMATION:
; APPLICANT: Vehmaaper, Jari
; APPLICANT: M nyl, Arja
; APPLICANT: Fagerstr m, Richard
; APPLICANT: Lantto, Raija
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko
; APPLICANT: Lahtinen, Taina
; APPLICANT: Kristo, Paula
; TITLE OF INVENTION: Actinomadura Xylanase Sequences and Methods
; TITLE OF INVENTION: of Use
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLSTEIN & FOX
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,812
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/332,412
FILING DATE: 31-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/282,001
FILING DATE: 29-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Larry B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 1050.0340002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540
TELEFAX: 202-371-2600
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 491 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
POSITION IN GENOME:
; CHROMOSOME/SEGMENT: M64551
US-08-468-812-8

Query Match 1.0%; Score 7; DB 2; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 LNSDGSV 211
Db 447 LNSDGSV 453

RESULT 22
US-08-590-563-8
; Sequence 8, Application US/08590563
; Patent No. 6300114
; GENERAL INFORMATION:
; APPLICANT: M nyl, Arja
; APPLICANT: Vehmaaper, Jari
; APPLICANT: Fagerstr m, Richard
; APPLICANT: Lantto, Raija

; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko
; APPLICANT: Lantinen, Tarja
; TITLE OF INVENTION: Production and Secretion of Proteins of
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W. Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/590,563
; FILING DATE: 26-JAN-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/468,812
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/332,412
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/282,001
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugaisky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 1050.0340003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 491 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: M64551
; US-08-590-563-8

Query Match 1.0%; Score 7; DB 4; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 LNSDGSV 211
Db 447 LNSDGSV 453

RESULT 23
US-09-319-989-6
; Sequence 6, Application US/09319989
; Patent No. 6190914
; GENERAL INFORMATION:
; APPLICANT: Grivell, Leslie A.
; APPLICANT: Teixeira De Mattos, Maarten J.
; APPLICANT: Blom, Jolanda
; TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
; TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
; TITLE OF INVENTION: METHODS
; FILE REFERENCE: 24615-20123.00
; CURRENT APPLICATION NUMBER: US/09/319,989

; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: PCT/NL97/00688
; EARLIER FILING DATE: 1997-12-12
; EARLIER APPLICATION NUMBER: EPO 96203520
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HAP4
; US-09-319-989-6

Query Match 1.0%; Score 7; DB 4; Length 554;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 498 YDLNDNN 504
Db 518 YDLNDNN 524

RESULT 24
US-09-134-001C-5569
; Sequence 5569, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5569
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5569

Query Match 1.0%; Score 7; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 KNGDYSS 57
Db 44 KNGDYSS 50

RESULT 25
US-08-656-177A-2
; Sequence 2, Application US/08656177A
; Patent No. 5882851
; GENERAL INFORMATION:
; APPLICANT: Koch, Birgit M.
; APPLICANT: Sibbesen, Ole
; APPLICANT: Halkier, Barbara Ann
; APPLICANT: Moller, Birger L.
; TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5862851artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/656,177A
APPLICATION NUMBER: US/08/656,177A
FILING DATE: 08-AUG-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19783/A/PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sorghum bicolor
IMMEDIATE SOURCE:
CLONE: P-450-Tyr
US-08-656-177A-2

Query Match 1.0%; Score 7; DB 2; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLA 20
|||||
Db 10 AATVLA 16

RESULT 26
US-09-256-797-2
Sequence 2, Application US/09256797
Patent No. 6133417
GENERAL INFORMATION:
APPLICANT: Koch, Birgit M.
APPLICANT: Sibbesen, Ole
APPLICANT: Halkier, Barbara Ann
APPLICANT: Moller, Birger L.
TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6133417artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/256,797
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/656,177
FILING DATE: 08-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19783/A/PCT

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sorghum bicolor
IMMEDIATE SOURCE:
CLONE: P-450-Tyr
US-09-256-797-2

Query Match 1.0%; Score 7; DB 4; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLA 20
|||||
Db 10 AATVLA 16

RESULT 27
US-09-564-805-240
Sequence 240, Application US/09564805
Patent No. 6333403
GENERAL INFORMATION:
APPLICANT: Tavtigian, Sean V.
APPLICANT: Teng, David H.F.
APPLICANT: Simard, Jacques
APPLICANT: Rommens, Johanna M.
TITLE OF INVENTION: Myriad Genetics, Inc.
TITLE OF INVENTION: Gene and a Paralog and Orthologous Genes
FILE REFERENCE: 2318-258
CURRENT APPLICATION NUMBER: US/09/564,805
CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/107,468
PRIOR FILING DATE: 1998-11-06
PRIOR APPLICATION NUMBER: 09/434,382
PRIOR FILING DATE: 1999-11-05
NUMBER OF SEQ ID NOS: 240
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 240
LENGTH: 661
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-564-805-240

Query Match 1.0%; Score 7; DB 4; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 478 EGGTLHK 484
|||||
Db 27 EGGTLHK 33

RESULT 28
US-09-347-878-16
Sequence 16, Application US/09347878C
Patent No. 6376210
GENERAL INFORMATION:
APPLICANT: Yuan, Chong
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
FILE REFERENCE: 25885-1651
CURRENT APPLICATION NUMBER: US/09/347,878C
CURRENT FILING DATE: 1999-07-06
NUMBER OF SEQ ID NOS: 75

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-347-878-16

Query Match
Best Local Similarity 1.0%; Score 7; DB 4; Length 714;
Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 KIESGEG 479
DB 523 KIESGEG 529
|||||

US-09-405-728-2
Sequence 2, Application US/09405728
Patent No. 6391316
GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Rioux, Clement
APPLICANT: Schryvers, Anthony B.
TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
FILE REFERENCE: 9000-0049.20
CURRENT APPLICATION NUMBER: US/09/405,728
CURRENT FILING DATE: 1999-09-24
EARLIER APPLICATION NUMBER: US 09/267,749
EARLIER FILING DATE: 1999-03-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 971
TYPE: PRT
ORGANISM: Haemophilus somnus
US-09-405-728-2

Query Match
Best Local Similarity 1.0%; Score 7; DB 4; Length 971;
Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 AAGFGTH 192
DB 537 AAGFGTH 543
|||||

US-08-026-138E-3
Sequence 3, Application US/08026138E
Patent No. 5502166
GENERAL INFORMATION:
APPLICANT: Masayoshi MISHINA
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nishiohata Residence 1-107
STREET: 5214, Nishiohata-machi
CITY: Niigata-shi
STATE: Niigata-ken
COUNTRY: JAPAN
ZIP: 951
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS v.5
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/026,138E
FILING DATE: 26-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 39563/1992

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-347-878-16

Query Match
Best Local Similarity 1.0%; Score 7; DB 4; Length 714;
Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 KIESGEG 479
DB 523 KIESGEG 529
|||||

US-09-405-728-2
Sequence 2, Application US/09405728
Patent No. 6391316
GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Rioux, Clement
APPLICANT: Schryvers, Anthony B.
TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
FILE REFERENCE: 9000-0049.20
CURRENT APPLICATION NUMBER: US/09/405,728
CURRENT FILING DATE: 1999-09-24
EARLIER APPLICATION NUMBER: US 09/267,749
EARLIER FILING DATE: 1999-03-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 971
TYPE: PRT
ORGANISM: Haemophilus somnus
US-09-405-728-2

Query Match
Best Local Similarity 1.0%; Score 7; DB 4; Length 971;
Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 AAGFGTH 192
DB 537 AAGFGTH 543
|||||

US-08-026-138E-3
Sequence 3, Application US/08026138E
Patent No. 5502166
GENERAL INFORMATION:
APPLICANT: Masayoshi MISHINA
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nishiohata Residence 1-107
STREET: 5214, Nishiohata-machi
CITY: Niigata-shi
STATE: Niigata-ken
COUNTRY: JAPAN
ZIP: 951
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS v.5
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/026,138E
FILING DATE: 26-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 39563/1992

; FILING DATE: 26-FEB-1992
; APPLICATION NUMBER: JP 173155/1992
; FILING DATE: 30-JUN-1992
; APPLICATION NUMBER: JP 215017/1992
; FILING DATE: 12-AUG-1992
; APPLICATION NUMBER: JP 303878/1992
; FILING DATE: 13-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hamburg, C.Bruce
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-4551
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 986-2340
; TELEFAX: (212) 953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 amino acids
; TYPE: amino acid
; STRANDEDNESS: single strand
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: mouse
; TISSUE TYPE: brain
; PUBLICATION INFORMATION:
; AUTHORS: Masayoshi MISHINA
; TITLE: NOVEL PROTEINS AND GENES CODING THE SAME
; RELEVANT RESIDUES IN SEQ ID NO: 3: FROM 1 to 1239
US-08-026-138E-3

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 1239;
Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ADVSGSL 205
DB 907 ADVSGSL 913
|||||

US-08-687-399-7
Sequence 7, Application US/08687399
Patent No. 5528381
GENERAL INFORMATION:
APPLICANT: Toft, Annette H.
APPLICANT: Marcher, Dorthie
APPLICANT: Pedersen, Hanne H.
APPLICANT: Nilsson, Thomas E.
TITLE OF INVENTION: A Combined Desizing and Bleaching
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5928381o No. 5928381disk of No. 5928381th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,399
FILING DATE:
CLASSIFICATION: 008
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4127.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
```

TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1385 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-687-399-7

Query Match 1.0%; Score 7; DB 2; Length 1385;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 ARGGLAL 413
Db 1001 ARGGLAL 1007

RESULT 32
US-09-413-814-46
; Sequence 46, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; TITLE OF INVENTION: heteropolypeptide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 1544
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-46

Query Match 1.0%; Score 7; DB 4; Length 1544;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 RGRVMAQ 218
Db 673 RGRVMAQ 679

RESULT 33
US-08-462-467B-4
; Sequence 4, Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The Use of a BMP Protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East Miami River Road

CITY: Ross
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Herko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1618 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-462-467B-4

Query Match 1.0%; Score 7; DB 4; Length 1618;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LYARGGL 411
Db 580 LYARGGL 586

RESULT 34
US-08-462-467B-2
; Sequence 2, Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The Use of a BMP Protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East Miami River Road
; CITY: Ross
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Herko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-462-467B-2

Query Match          1.0%; Score 7; DB 4; Length 2887;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LYARGGL 411
DB 580 LYARGGL 586

RESULT 35
US-08-462-467B-8
; Sequence 8, Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The Use of a BMP Protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; TITLE OF INVENTION: BMP Receptor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East Miami River Road
; CITY: Ross
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hersko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-462-467B-8

Query Match          1.0%; Score 7; DB 4; Length 2887;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LYARGGL 411
DB 580 LYARGGL 586

RESULT 36
US-08-159-339A-836
; Sequence 836, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Kubo, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
```

```

; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 836:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-836

Query Match          0.8%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 RGPSSL 159
DB 1 RGPSSL 6

RESULT 37
US-08-159-339A-891
; Sequence 891, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Kubo, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
```


STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,339A
FILING DATE: 29-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,666
FILING DATE: 07-AUG-1992
APPLICATION NUMBER: US 08/027,746
FILING DATE: 05-MAR-1993
APPLICATION NUMBER: US 08/103,396
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-005030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX:
INFORMATION FOR SEQ ID NO: 891:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-891

Query Match 0.8%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 RGPSTL 159
|||||
Db 5 RGPSTL 10

RESULT 38
US-08-191-866D-64
Sequence 64, Application US/08191866D
Patent No. 5783195
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Macdonald, Richard D.
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus S-IBR-052 And Uses Thereof
NUMBER OF SEQUENCES: 99
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/191,866D
FILING DATE: 4 February 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.

REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-191-866D-64

Query Match 0.8%; Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 YKIESG 477
|||||
Db 2 YKIESG 7

RESULT 39
US-08-185-949B-64
Sequence 64, Application US/08185949B
Patent No. 5874279
GENERAL INFORMATION:
APPLICANT: Mark D. Cochran
APPLICANT: Richard D. Macdonald
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,949B
FILING DATE: 03-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 278-0525
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-185-949B-64

Query Match 0.8%; Score 6; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 YKIESG 477
|||||
Db 2 YKIESG 7

RESULT 40
PCT-US93-07306-45

Sequence 45, Application PC/TUS9307306
GENERAL INFORMATION:
APPLICANT: MARGOLIS, Richard U.
APPLICANT: RAUCH, Uwe
APPLICANT: MARGOLIS, Renee K.
TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Broadway and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07306
FILING DATE: 03-AUG-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,911
FILING DATE: 03-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Guy K.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: Margolis=1A PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-07306-45

Query Match 0.8%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 SDGSVR 212
Db 3 SDGSVR 8

RESULT 41
US-08-520-933-2
Sequence 2, Application US/08520933
Patent No. 5981194
GENERAL INFORMATION:
APPLICANT: Jefferies, Wilfred A.
APPLICANT: McGeer, Patrick L.
APPLICANT: Rothenberger, Sylvia
APPLICANT: Food, Michael R.
APPLICANT: Yamada, Tatsuo
APPLICANT: Kennard, Malcolm
TITLE OF INVENTION: Use of p97 and Iron Binding Proteins
as Diagnostic and Therapeutic Agents
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/520,933
FILING DATE: August 31, 1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Shona S. McDiarmid
REGISTRATION NUMBER: 38,798
REFERENCE/DOCKET NUMBER: 7685-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-364-7311
TELEFAX: 416-361-1398
TELEX: 06-23115
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-520-933-2

Query Match 0.8%; Score 6; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 MRGPSS 158
Db 1 MRGPSS 6

RESULT 42
US-09-285-040-2
Sequence 2, Application US/09285040
Patent No. 6455494
GENERAL INFORMATION:
APPLICANT: Jefferies, Wilfred A.
APPLICANT: McGeer, Patrick L.
APPLICANT: Rothenberger, Sylvia
APPLICANT: Food, Michael R.
APPLICANT: Yamada, Tatsuo
APPLICANT: Kennard, Malcolm
TITLE OF INVENTION: Use of p97 and Iron Binding Proteins
as Diagnostic and Therapeutic Agents
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/285,040
FILING DATE: 01-APR-1999
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7685-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-364-7311
TELEFAX: 416-361-1398
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-285-040-2

Query Match 0.8%; Score 6; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 MRPSG 158
Db 1 MRPSG 6

RESULT 43
US-08-256-747C-59
Sequence 59, Application US/08256747C
Patent No.6037448
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: GILLAM, Shirley
APPLICANT: OU, Dawei
APPLICANT: TINGE, Aubrey
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR A RUBELLA VACCINE
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,747C
FILING DATE: 06-OCT-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-370
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-256-747C-59

Query Match 0.8%; Score 6; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 SSSVFA 27
Db 2 SSSVFA 7

RESULT 44
US-08-834-130A-59
Sequence 59, Application US/08834130A
Patent No. 6180758
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: GILLAM, Shirley
APPLICANT: OU, Dawei

APPLICANT: TINGE, Aubrey
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR A RUBELLA VACCINE
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,130A
FILING DATE: 14-APR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-686 MIS:jib
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-834-130A-59

Query Match 0.8%; Score 6; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 SSSVFA 27
Db 2 SSSVFA 7

RESULT 45
US-08-726-306A-118
Sequence 118, Application US/08726306A
Patent No. 5958684
GENERAL INFORMATION:
APPLICANT: van Leeuwen, Frederik Willem
APPLICANT: Burbach, Johannes Peter Henri
APPLICANT: Grosveld, Franklin G.
TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
NUMBER OF SEQUENCES: 189
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1 Financial Center
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: wordperfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,306A
FILING DATE: 02-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 95/20080.4
FILING DATE: 02-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/009,832

;
; FILING DATE: 01-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-726-306A-118

Query Match 0.8%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 SPRPAE 228
|||||
Db 4 SPRPAE 9

Search completed: December 18, 2002, 07:00:10
Job time : 30.7742 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:38:19 ; Search time 49.0531 Seconds
(without alignments)
3032.756 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776
Sequence: 1 MGQFNSVFRINMTAATVLA.....NTFNIGSERITWNLIRYSF 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3244	85.9	708	16	Q9JXZ0
2	832	22.0	802	16	Q9HML3
3	821.5	21.8	815	2	Q51339
4	788	20.9	725	16	Q9JW26
5	784	20.8	725	16	Q9JXU3
6	755	20.0	753	16	Q8EFY6
7	736.5	19.5	729	16	Q8X8H4
8	721	19.1	724	16	Q8ZQ10
9	656	17.4	734	2	Q9X6A5
10	512	13.6	863	16	Q8YV33
11	486.5	12.9	820	16	Q8YV33
12	486.5	12.9	885	16	Q8YV33
13	483.5	12.8	854	16	Q8YV33
14	481	12.7	858	16	Q8YV33
15	479	12.6	851	16	Q8YV33
16	475	12.6	863	16	Q8YV06

17	471.5	12.5	733	16	Q926C7	Q926C7 rhizobium m
18	467.5	12.4	867	16	Q8YV33	Q8YV33 anabaena sp
19	465	12.3	828	16	P72602	P72602 synecocyst
20	463	12.3	695	16	Q8XV06	Q8XV06 ralsstoma s
21	460.5	12.2	851	16	Q8YV33	Q8YV33 anabaena sp
22	460	12.2	760	16	Q8XV33	Q8XV33 escherichia
23	459	12.2	708	16	Q9HUX3	Q9HUX3 pseudomonas
24	458	12.1	726	2	Q9X7K9	Q9X7K9 rhizobium 1
25	458	12.1	747	2	Q9JUP3	Q9JUP3 rhizobium 1
26	450.5	11.9	747	16	Q8U1B9	Q8U1B9 agrobacteri
27	450	11.9	819	16	Q8UK13	Q8UK13 agrobacteri
28	449.5	11.9	872	16	Q8YV33	Q8YV33 anabaena sp
29	441.5	11.7	753	16	Q8E590	Q8E590 pseudomonas
30	436.5	11.6	732	2	Q6E599	Q6E599 pseudomonas
31	432.5	11.5	732	16	Q91422	Q91422 pseudomonas
32	432	11.4	755	2	Q9F0F9	Q9F0F9 campylobact
33	429	11.4	732	2	Q8E424	Q8E424 pantoea agg
34	424	11.2	858	16	Q8YV33	Q8YV33 anabaena sp
35	423	11.2	724	16	Q926C6	Q926C6 rhizobium m
36	421.5	11.2	696	16	Q82908	Q82908 salmoneilla
37	420.5	11.1	853	16	Q8YV33	Q8YV33 anabaena sp
38	416.5	11.0	853	16	P72609	P72609 synecocyst
39	415.5	11.0	857	16	Q8YV33	Q8YV33 anabaena sp
40	412	10.9	689	16	Q8XV08	Q8XV08 ralsstoma s
41	411	10.9	714	16	Q92YX1	Q92YX1 rhizobium m
42	410.5	10.9	762	2	Q9E293	Q9E293 zymomonas m
43	410	10.9	801	16	Q8XV31	Q8XV31 ralsstoma s
44	407.5	10.8	635	16	Q9JTR6	Q9JTR6 neisseria m
45	402	10.6	863	16	P72599	P72599 synecocyst

ALIGNMENTS

RESULT 1

Q9JXZ0 PRELIMINARY; PRT; 708 AA.
AC Q9JXZ0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE TonB-dependent receptor.
GN NMB1829.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Kouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masigam V., Pizzano R., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappunoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58.";
RL Science 287:1809-1815 (2000).
RT EMBL; AE002532; AAF42164.1; -.
DR TIGR; NMB1829; -.
DR InterPro; IPR002106; AATRNA_LigaseII.
DR InterPro; IPR001917; NHRtransf_2.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00599; AA_TRANSF_CLASS_2; UNKNOWN 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN 1.
KW Receptor; Complete Proteome;
SQ SEQUENCE 708 AA; 77068 MW; F3519B024FBB4EE CRC64;

Query Match 85.9%; Score 3244; DB 16; Length 708;

Best Local Similarity 86.8%; Pred. No. 1.4e-182;
Matches 627; Conservative 33; Mismatches 48; Indels 14; Gaps 3;

QY 1 MQQFMSVFRIINTAATVLAALSSVFAAQATABLETHVHKGQSYNAIVTEKNGDYSSFAV 60
DB 1 MQQFMSVFRIINTAATVLAALSSVFAAQATABLETHVHKGQSYNAIVTEKNGDYSSFAA 60

QY 61 TVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQARKTGPLRLVLSNDGGRSSVYARG 120
DB 61 TVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQARKTGPLRLVLSNDGGRSSVYARG 120

QY 121 EYSEYNIDGLPAQMOSINGTLNLFADFVEVMRGPGLFDSSGEMGGIVNLRKPTKA 180
DB 121 EYSEYNIDGLPAQMOSINGTLNLFADFVEVMRGPGLFDSSGEMGGIVNLRKPTKA 180

QY 181 FOGHAAAGFTGHQYKAEADVSGSLNSDCSVGRVMAQTVGASPRPAEKNNRHETFYAAA 240
DB 181 FOGHAAAGFTGHQYKAEADVSGSLNSDCSVGRVMAQTVGASPRPAEKNNRHETFYAAA 240

QY 241 DWDINDPTVLGAGLYQOORHLAPYNGLPADANNKPLSLPQHVFVGADNMKFKQNSHDVFA 300
DB 241 DWDINDPTVLGAGLYQOORHLAPYNGLPADANNKPLSLPQHVFVGADNMKFKQNSHDVFA 300

QY 301 DLKHYFGNGGKGVGNRYSDRDADSNYAFAGSKLGKMTAGRPAGCGNTADDKACAVGLGTE 360
DB 301 DLKHYFGNGGKGVGNRYSDRDADSNYAFAGSKLGKMTAGRPAGCGNTADDKACAVGLGTE 360

QY 361 IKOKALAFDASYSRPPRLGNTANEFVIGADYNFRSTNEQGRSTLYARGGLALNEFRSIP 420
DB 361 IKOKALAFDASYSRPPRLGNTANEFVIGADYNFRSTNEQGRSTLYARGGLALNEFRSIP 420

QY 421 QVDLIANARKVGRGYSHTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGCK 480
DB 421 QVDLIANARKVGRGYSHTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGCK 480

QY 481 TLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVGK 540
DB 481 TLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVGK 540

QY 541 GSYMDRLNARYSFYRMKDKNAAAPLNPNKTRVAALGKRWMEGVETEISGAVTPKQI 600
DB 541 GSYMDRLNARYSFYRMKDKNAAAPLNPNKTRVAALGKRWMEGVETEISGAVTPKQI 600

QY 601 HAGYSLHSQIKTASNRDGLIFLLMPKHSANLWTTYQVTPBLTITGGGVNAMSGITSSAG 660
DB 601 HAGYSLHSQIKTASNRDGLIFLLMPKHSANLWTTYQVTPBLTITGGGVNAMSGITSSAG 660

QY 661 MHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGGANTENIPGSERTWTANLRY 720
DB 661 MHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGGANTENIPGSERTWTANLRY 720

QY 721 SF 722
DB 707 SF 708

RESULT 2

Q9HWL3 PRELIMINARY; PRT; 802 AA.
AC Q9HWL3
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE Probable TonB-dependent receptor.
GN PA4168.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NCBI TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004833; AAG07555.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 802 AA; 87430 MW; 379963AE03B47EAA CRC64;

Query Match 22.0%; Score 832; DB 16; Length 802;

Best Local Similarity 30.0%; Pred. No. 9.6e-41;

Matches 222; Conservative 137; Mismatches 309; Indels 72; Gaps 21;

QY 19 AALSSVFAAQATABLETHVHKG-QRSYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQSV 77
DB 19 AALSSVFAAQATABLETHVHKG-QRSYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQSV 77

QY 97 AATGAGIAPGSDVLGALSVVATQDPATAAITEDSGSYTTRAMRTSTKLAMSIRETQSV 156
DB 97 AATGAGIAPGSDVLGALSVVATQDPATAAITEDSGSYTTRAMRTSTKLAMSIRETQSV 156

QY 78 SIITNQVKDRNVDTFDQARKTGPLRLVLSNDGGRSSVYARGYSEYNIDGLPAQMOSI 137
DB 78 SIITNQVKDRNVDTFDQARKTGPLRLVLSNDGGRSSVYARGYSEYNIDGLPAQMOSI 137

QY 157 SVVTRQMRDGDGMDLNDADVKGVTGLTVQYGPARYAARGFVDNIMYDGLPTSTISY 216
DB 157 SVVTRQMRDGDGMDLNDADVKGVTGLTVQYGPARYAARGFVDNIMYDGLPTSTISY 216

QY 138 NG---TLNLPADFPRVEVMRGPGLFDSSGEMGGIVNLRKPTKAFQGHAAAGFTGHQ 194
DB 138 NG---TLNLPADFPRVEVMRGPGLFDSSGEMGGIVNLRKPTKAFQGHAAAGFTGHQ 194

QY 217 TQDVISAADLAFDFVEVVRGATGLMQAGNPAAAINNVKRPQTEFRASLOGSVGTWDR 276
DB 217 TQDVISAADLAFDFVEVVRGATGLMQAGNPAAAINNVKRPQTEFRASLOGSVGTWDR 276

QY 195 YKABADVSGSLNSDCSVGR-VMAQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAG 253
DB 195 YKABADVSGSLNSDCSVGR-VMAQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAG 253

QY 277 YRSEADVSGPLNSEGLRGRAYVAYQKNSPDHVSSE-GVFYGIAEADLSPDTTFAIG 335
DB 277 YRSEADVSGPLNSEGLRGRAYVAYQKNSPDHVSSE-GVFYGIAEADLSPDTTFAIG 335

QY 254 YLYQORHLAP-YNGLPADANNKPLSLPQHVFVGADNMKFKQNSHDVFPADLKHFGNGGYG 312
DB 254 YLYQORHLAP-YNGLPADANNKPLSLPQHVFVGADNMKFKQNSHDVFPADLKHFGNGGYG 312

QY 336 ASNQNDNRNDNMVGLPGLGRRHLDLKRSSYYGADWSYMDTDTTHLFSDLTHRFANGQM 395
DB 336 ASNQNDNRNDNMVGLPGLGRRHLDLKRSSYYGADWSYMDTDTTHLFSDLTHRFANGQM 395

QY 313 KVMRYSDRDADSNYAFAG-----SKLGKMTAGRP-CNTADDKACAVGLGTEI 361
DB 313 KVMRYSDRDADSNYAFAG-----SKLGKMTAGRP-CNTADDKACAVGLGTEI 361

QY 396 KLA-----ADKLWARINMLGLYNDCYSTTGCASMTQNPQDYSYTDDHD----- 439
DB 396 KLA-----ADKLWARINMLGLYNDCYSTTGCASMTQNPQDYSYTDDHD----- 439

QY 362 KOKALAFDASYSRPPRLGNTANEFVIGADYNFRSTNEQGRSTLYARGGL---ALNEFRS 418
DB 362 KOKALAFDASYSRPPRLGNTANEFVIGADYNFRSTNEQGRSTLYARGGL---ALNEFRS 418

QY 440 ----SYDAYANGPQLLGRHEHLVWGSYRQERFDGHHGWSLFNKDGCTPTGPMPTQW 494
DB 440 ----SYDAYANGPQLLGRHEHLVWGSYRQERFDGHHGWSLFNKDGCTPTGPMPTQW 494

QY 419 IPOVDLIANARKVGRGYSHTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGE 478
DB 419 IPOVDLIANARKVGRGYSHTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGE 478

QY 495 DPSSTLKPRLWNTSLWGMK-----LDQEQKGAYLTTRLNADPLKVLGGRLDWDYKADADT 549
DB 495 DPSSTLKPRLWNTSLWGMK-----LDQEQKGAYLTTRLNADPLKVLGGRLDWDYKADADT 549

QY 479 GKTLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVG 538
DB 479 GKTLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVG 538

QY 550 DS--YKTR-NVTRYAGVIYDLNQTYSYASYTDIIFKQSNFDAGGGLDITGKNYEIG 606
DB 550 DS--YKTR-NVTRYAGVIYDLNQTYSYASYTDIIFKQSNFDAGGGLDITGKNYEIG 606

QY 539 YGYSMDRLNARYSFYRMKDKNAA-----APLNPNKTRVA--ALGKRWMEGV 587
DB 539 YGYSMDRLNARYSFYRMKDKNAA-----APLNPNKTRVA--ALGKRWMEGV 587

QY 607 LKGEHFGGALNSQIALFOIDQENRATEDVGGPSCPFTSPT---SRYCSRASGKVRSGVD 663
DB 607 LKGEHFGGALNSQIALFOIDQENRATEDVGGPSCPFTSPT---SRYCSRASGKVRSGVD 663

QY 588 TEISGAVTPKQIHAAGYSLHSQIKTASNRDGL--IFLLMPKHSANLWTTYQVTPEL-- 643
DB 588 TEISGAVTPKQIHAAGYSLHSQIKTASNRDGL--IFLLMPKHSANLWTTYQVTPEL-- 643

QY 664 LELSGALSDDDQMOMAGYTYDAKYKHSNKANEGKPFDAKPRHLFKLATSYTLPGEIHK 723
DB 664 LELSGALSDDDQMOMAGYTYDAKYKHSNKANEGKPFDAKPRHLFKLATSYTLPGEIHK 723

QY 644 -TIGGVNAMSIGI-TSSAGMHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGG 701
DB 644 -TIGGVNAMSIGI-TSSAGMHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGG 701

QY 724 WRVGGDLATQSKTEDSSTGFGQGGYTVVNMGLKYKNERIDTRLNFNLLFKYIYSGIDF 783
DB 724 WRVGGDLATQSKTEDSSTGFGQGGYTVVNMGLKYKNERIDTRLNFNLLFKYIYSGIDF 783

QY 702 ANTFNIPGSERTWTANLRY 721
DB 702 ANTFNIPGSERTWTANLRY 721

QY 784 GN-LNY-GEPRNLMFTVKYS 801
DB 784 GN-LNY-GEPRNLMFTVKYS 801

RESULT 3

```
051339
ID 051339 PRELIMINARY; PRT; 815 AA.
AC 051339;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN Ferrityoverdine receptor.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_Taxid=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=93328663; PubMed=8335619;
RA Poole K., Neshat S., Krebes K., Heinrichs D.B.;
RT "Cloning and nucleotide sequence analysis of the ferrityoverdine
RT receptor gene fpya of Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:4597-4604(1993).
RN [2]
RP SEQUENCE OF 552-815 FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=95095977; PubMed=7798141;
RA Merriman T.R., Merriman M.E., Lamont I.L.;
RT "Nucleotide sequence of pvdD, a pyoverdine biosynthetic gene from
RT Pseudomonas aeruginosa: PvdD has similarity to peptidic synthetases.";
RL J. Bacteriol. 177:252-258(1995).
RN [3]
RP SEQUENCE OF 1-551 FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97075909; PubMed=8918232;
RA McMoran B.J., Merriman M.E., Rombel I.T., Lamont I.L.;
RT "Characterisation of the pvdB gene which is required for pyoverdine
RT synthesis in Pseudomonas aeruginosa.";
RL Gene 176:55-59(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RA Lamont I.L.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U07359; AAB60199.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KM Receptor.
FT CONFLICT 750 750 S -> R (IN REF. 1).
FT CONFLICT 716 716 Y -> F (IN REF. 1).
SQ SEQUENCE 815 AA; 91167 MW; A11AE4A5A290F35 CRC64;

Query Match 21.8%; Score 821.5; DB 2; Length 815;
Best Local Similarity 29.0%; Pred. No. 4.1e-40;
Matches 217; Conservative 145; Mismatches 312; Indels 75; Gaps 24;

QY 8 FRINMTAATVLAALSSVFAAQTADLETVAHIKQGRSYNAIVTEKNGDYSSFAVTVGTGIP 67
DB 108 FQGNATITISVAEADSSV-----DLGATWITSNQL--GTTEDSGSYTGTIATATRLV 159
QY 68 ASARELPQSVITITNOQVAKRNDVTPQGLARKTPGLVLSNDGRSSVYARGEYSYNI 127
DB 160 LTPRETPQSVITVTRQMDPFGNNIDVVRHHTPGIVSAVYDDRNRYVARGSSINNFQY 219
QY 128 DGLPAQMOSI---NGTLPLVLPFADRYEVMRSPGLPDSGEGMGIVNLVRKPTAKFOG 163
DB 220 DGLPSTARNVGYAGNTLSDMALYDRVEVLKATGLTSGAGATNLLRKKEPTHEFGK 279
QY 184 HAAAGFTHKQYKAADVSGSLNSDGSVGRVVAQTVGASPPAKKNRHETFYAADM 243
DB 280 HVELGAGSMQNVYSELDSVGLPTESGVNRGAVAAVQDKSPMDHYERKTSVYVYGLIEFD 339
QY 244 INPDYLVAGAYLQQRHL--APYNG-LPA--DANNKLPSLFQHYFVGDANKKFKMNSHDVF 299
DB 340 LNDPTLVGADYQDDNPKSGWSGSPFLPDSQGNRNDVSRFNGAKWSSWEOYTRTVF 399
```

```
QY 300 ADLKHVFNNGYGVKGYMRYSRDRDASNVAFAAGSKLGKTPAGRPGCNTADDKACAVGLGT 359
DB 400 ANLEHNPANGWGVKQL----DHKINGYHAPLGLIMG-DWPA-----PDNSAKIVAQKY 448
QY 360 EIQKALAPFASYSRPPRLGNTANFVIG--ADYNRFRSTNEGRTTYLARG--GLALNEF 416
DB 449 TGETKSNSLDIYLTGPPQFGRHEHLVVGTSASFHW-----EGKSYMLRNNDNTTDDF 503
QY 417 RSLPQVDLLANARKRGVSHVATENLDFEFGYGSTFHPADGLSLGGRLGHYKIES 476
DB 504 -----INMGDICKPDMGTPSQYIIDKTRQLGSGYMTAREVNTDNLNPLGGRVVDTRV-T 557
QY 477 GEGKTLHKASKTKFTGYAGAVYDLDNNSLYLSLSQLYTPQTN--LDADGKLKPRQNG 534
DB 558 GLNPRTRESG--RFIYVGAVYDLDNTYSVYASYTDIIMPQDSWYDSSKLLPDEGQN 615
QY 535 FEVYGYGSTMDDPLNARVSYRMRKDKNAAP-----LNPNNKTRYAALG--KRVMEGVET 588
DB 616 YEIGIKGEYLDGRLNTSLAVFEIHERRAEDALVNSKPTNPATYAAVKGIKAKTKGYEA 675
QY 589 EISGAVTPKQIHAQSYLSHSQKTASNSRDGIFILMPGHSANLMTTQO--VTBELTI 645
DB 676 EISGELAPGMQVQAGYT--HKIIRDSGKK--VSTWEPQDLSLYTSYFKGALDKLTV 730
QY 646 GGGV-----NMSGITSSAGMHAGYATPDMAAAYFTPKLKLQINADNI FNR 693
DB 731 GGGARRQKSGMWQVNNPNRSMK---FSGEDYVYLDIARVYITDKLSASVAVNNVFDK 787
QY 694 HYYARVGANTFNIPGSEERTANLRYSF 722
DB 788 TYTYNTIGFYTSASY-GDPRNIMPSTRWDF 815

RESULT 4
QY 05JW26 PRELIMINARY; PRT; 725 AA.
AC 05JW26;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative ferric siderophore receptor protein.
GN MAA0575.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OC NCBI_Taxid=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
RA Jagers K., Leather S., Moulé S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL; AL162753; CAB83866.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
KM Receptor; Complete proteome.
SQ SEQUENCE 725 AA; 80302 MW; DDBBF76A024EPAB6 CRC64;

Query Match 20.9%; Score 788; DB 16; Length 725;
Best Local Similarity 29.9%; Pred. No. 3.2e-38;
Matches 232; Conservative 117; Mismatches 316; Indels 112; Gaps 23;

QY 5 MSVFRINMTAATVLAALSSVFAAQTADLETVAHIKQGRSYNAIVTEKNGDYSS 57
DB 1 MTRFKYSLFPAALLPYVAQADVSDDPKQESTELPITTVTADRT-----ASSNDGYTV 55
```

```

QY 58 FAVTVGKIPASLREIPOSVSIIITNOQVKDRNVDTFQ--LARKTPGLRLVLSNDD--GRSS 114
Db 56 SGTHPTPLGLPMTLREIPOSVSITSQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNY 115
QY 115 VYARGVEYSEYNIDGLPAQMOSINGTLPNLFADPDRVEMRGPSGLFDSSGEMGGIVNLVR 174
Db 116 LFARGSRANYQINGIPVADALADTGNANTAAVERVEVRGVAGLLDGTGEPSTVNLVR 175
QY 175 KRPTKAFQGHAAAGFTHKQYKAEADVSGSLNSDGSVRGRVMAQ--TVGASPRPAEKNRHH 233
Db 176 KRPTKPLFEVRAEAGNRKHFLGADVSGSLNAEGTLRGLVSTFGRGDSMRQRER--SRD 234
QY 234 EFTYAADWDINDPDTVLGAGLYQQRHLL---APYNGLPADANNKLPSPLOHVFGVADWKNF 290
Db 235 AELYGILEYDIAPOTRVHAGMDYQAKETADAPLSYAVYDSQYATAFGPKDNPATNMAN 294
QY 291 FQWNSHDVFDLKHVFGNGGKVGWRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADD 350
Db 295 SHRALNLPAGIEHRFN-----QDWKLKAEYDITRSR--FRQPYGVAGVLSIDH 341
QY 351 KACAVGL---GTEIKOKALAFDASYSRPERLGNANTANEFVIGADYNRFRSTNEQGRITLYA 407
Db 342 NTAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIAGINGYKAS--NKYGE----- 395
QY 408 RGLALNFRSPQVDLLANARKVGRGYSHTVA-----TENLDEFGI-----YKST 454
Db 396 -----RST-----IPNAIPNAYEFSRTGAYPOPASFAQTIPQYGTTRQIGYLATR 441
QY 455 FHPADGLSLIGGRLGHYKIESGEGKT--LHKASKTKFTGYAGAVYDLNDNNSLYLSLQ 512
Db 442 FRAADNLSLILGGRYRTRTGSYDSRTQGTMTVYSANRFTPTGTGIVFDLTGNLSLGSYSS 501
QY 513 LVTPTQNLADGKLLKPRQGNQFEVYKGSYMDRLNARVSYRMDKN--AAAPLNPNK 570
Db 502 LFVPSQKDEHGSYLKPVTTGNLLEAGIKGEWLEGRNLNAAVYRARKNNLATAAGRDPSG 561
QY 571 KXTRVAALGKRWEGVETEISGAVTPKQIHAQYSLHSQIKTASNRDDGIFLMPKHS 630
Db 562 -NTYYRAANQAOKTHGWEIEVGGRIITPEWQIQAGYSQSKTRDQDGSRLNPDSV-----PERS 616
QY 631 ANLWTTYQVTPEL---TIGGGV-----NAMSGITSSAGMHAGGYATFD 670
Db 617 FKLFTAYHFAPAPSGWTIGAGVRQWSETHDTPATLRIPNPAKARAADNSRQKAYAVAD 676
QY 671 AMAAYRFTPKLKLQINADNIFNRHYARVGGANTFNIP-----GSERTWTANLRYSF 722
Db 677 IMARYRFPNRAELSLNVDNLFNKHYRTQ-----PDRHSYGALRTVNAAFYVRF 724

RESULT 5
Q9JXU3 PRELIMINARY; PRT; 725 AA.
AC Q9JXU3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE TonB-dependent receptor.
GN NMB1882.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza G., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;

```

```

RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RL MC58."; 287:1809-1815(2000).
DR EMBL; AE002538; AAF42216.1; -.
DR TIGR; NMB1882; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR Receptor; Complete proteome.
SQ SEQUENCE 725 AA; 80478 MW; D3953D4485FD8FAF CRC64;

Query Match 20.8%; Score 784; DB 16; Length 725;
Best Local Similarity 29.4%; Pred. No. 5 Se-38;
Matches 228; Conservative 118; Mismatches 320; Indels 110; Gaps 21;

QY 5 MSVFRINNTAATVL-----AALSSSVFAAQATADLETVHIKQRSYNAINVTEKNGDYSS 57
Db 1 MTRFKYSLFAALLPVYAQADVSDDPKQESTELPTITVTDRT-----ASSNDGYTV 55
QY 58 FAVTVGKIPASLREIPOSVSIIITNOQVKDRNVDTFQ--LARKTPGLRLVLSNDD--GRSS 114
Db 56 SGTHPTPLGLPMTLREIPOSVSITSQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNY 115
QY 115 VYARGVEYSEYNIDGLPAQMOSINGTLPNLFADPDRVEMRGPSGLFDSSGEMGGIVNLVR 174
Db 116 LFARGSRANYQINGIPVADALADTGNANTAAVERVEVRGVAGLLDGTGEPSTVNLVR 175
QY 175 KRPTKAFQGHAAAGFTHKQYKAEADVSGSLNSDGSVRGRVMAQTVGASPRPAEKNRHH 234
Db 176 KRPTKPLFEVRAEAGNRKHFLGADVSGSLNTEGTLRGLVSTFGRGDSWRRERSRDA 235
QY 235 TFYAAADWDINDPDTVLGAGLYQQRHLL---APYNGLPADANNKLPSPLOHVFGVADWKNF 291
Db 236 ELYGILEYDIAPOTRVHAGMDYQAKETADAPLSYAVYDSQYATAFGPKDNPATNMAN 295
QY 292 KQWNSHDVFDLKHVFGNGGKVGWRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDK 351
Db 296 SHRALNLPAGIEHRFN-----QDWKLKAEYDITRSR--FRQPYGVAGVLSIDHN 342
QY 352 KACAVGL---GTEIKOKALAFDASYSRPERLGNANTANEFVIGADYNRFRSTNEQGRITLYA 408
Db 343 TAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIAGINGYKAS--NKYGE----- 395
QY 409 GGLALNFRSPQVDLLANARKVGRGYSHTVA-----TENLDEFGI-----YKSTP 455
Db 396 -----RST-----IPNAIPNAYEFSRTGAYPOPASFAQTIPQYGTTRQIGYLATR 442
QY 456 HPADGLSLIGGRLGHYKIESGEGKT--LHKASKTKFTGYAGAVYDLNDNNSLYLSLQ 513
Db 443 RAADNLSLILGGRYRTRTGSYDSRTQGTMTVYSANRFTPTGTGIVFDLTGNLSLGSYSS 502
QY 514 YTPQTNLDADGKLLKPRQGNQFEVYKGSYMDRLNARVSYRMDKN--AAAPLNPNK 571
Db 503 FVPSQKDEHGSYLKPVTTGNLLEAGIKGEWLEGRNLNAAVYRARKNNLATAAGRDPSG- 561
QY 572 KTRVAALGKRWEGVETEISGAVTPKQIHAQYSLHSQIKTASNRDDGIFLMPKHS 631
Db 562 NTTYRAANQAOKTHGWEIEVGGRIITPEWQIQAGYSQSKTRDQDGSRLNPDSV-----PERS 617
QY 632 NLWTTYQVTPEL---TIGGGV-----NAMSGITSSAGMHAGGYATFDDA 671
Db 618 KLFYAYHFAPAPSGWTIGAGVRQWSETHDTPATLRIPNPAKARAADNSRQKAYAVADI 677
QY 672 MAAYRFTPKLKLQINADNIFNRHYARVGGANTFNIP-----GSERTWTANLRYSF 722
Db 678 MARYRFPNRAELSLNVDNLFNKHYRTQ-----PDRHSYGALRTVNAAFYVRF 724

RESULT 6
Q8ZFYG PRELIMINARY; PRT; 753 AA.
ID Q8ZFYG
AC Q8ZFYG;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

```



```

Db 127 IDNYVDGIPYFSSRNWLDALSDMALFERVEVVRGATGLTGTGNPSAIAINVRKHAT 186
Qy 179 -KAFQGHAAAGFGTHQYKABADVSGSLNSGSRVGRVMAQTGVA---SPRPAEKNNRHE 234
Db 187 SREFKGDVSAEYGSWNKERYVADLQSLPTEDGKIRARI---VGGYQNWDSWLDRYANSEK 242
Qy 235 TFYAA-ADWDINDPTVLGAGLYLQQRHL--APYNGLPA-DANNKLPSPQHVFGADWNK 290
Db 243 TFFSGIVDADLGDLTTLTLAGYEQYRIDVNSPTWGLPRWNTDGSNSYSDRARSTAPDWAY 302
Qy 291 FKMNSHDVADLKHYFGNGGKVGCMRYSRDDADSNY---AFAGSKLGMKTP----- 339
Db 303 NDKELNKVFWTLKQRPADTQWATLNATHSEVEFDSKMYVDAYVYNKADGMLVGYNSYGP 362
Qy 340 ----AGRPGCNTADDKACAVGLGTE-----IKQALAFDASYSRPFR--LGNANEF-- 385
Db 363 GFDYVGGTGMNSGKRVKVDALDLFADGSYELFGRQHNLNMFSGYSKQNNRYFSSWANIPPD 422
Qy 386 VIGADYNFRSTNEQRTILYARGGLALNEFRSIPQVDLIANARKGVGYSHTVATENLD 445
Db 423 EIGSPYN-FNG-----NFPQTD-----WSPOSQAQDDTT 450
Qy 446 EF-GYIGKSTFHPADGLSLIGGRGLGHYKIESGEGKTLHKASKTKFTGYAGAVDLDNDN 504
Db 451 HMKSLYAAATRVTLADPLHLILGARYTNVRDT----LTYSMEKNHTTYPYAGLVFINDNW 506
Qy 505 SLYLSLSQLYTPQTNLDADGKLLKPROGQFEGVYKGSYMDRLNARVSFYRMKDKNA-- 563
Db 507 STYASVTSIFQPNDRDSSGKYLAIPITGNVYELGLKSDWMSRLTTLTAIFRIEQDNVAQ 566
Qy 564 ---APLNPNKKTRYAALGKRWEGVEISGAVTPKQIHHAGYS-YLHQSQIKTASNRD 619
Db 567 STGTPIPSNGSETAYKAVDGTYSKGVPELNGAIDNNQLTFGATRYI-----AEDNEG 620
Qy 620 DGIFLLMPKHSANLWTTQ--VTPLETTGGVNMMSGITSSA-----GMHAGGYATFD 670
Db 621 NNVNPNLRTTVMKTSVRLPWPMLTGVGGVNWQNRVYTTVTPYGTFRAEQGSYALVD 680
Qy 671 AMAAYRFTPKLQINADNIFNRHYARVGGANTNIPGSERTWTANLRYSP 722
Db 681 LFTRYQVTKNFSLOGNVNLPDKYTDNVEGSIVY---GAPNFSITGTQYF 729

RESULT 8
Q8ZQ10 PRELIMINARY; PRT; 724 AA.
AC Q8ZQ10;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Out-er membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B
DE and Fe(III)-rhodotulic acid uptake.
GN FHUE OR STM1204.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmomella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
DR EMBL; AE008752; AAL20133.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.

```

```

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 724 AA; 80893 MW; 94622BFE3714BFC2 CRC64;

Query Match 19.1%; Score 721; DB 16; Length 724;
Best Local Similarity 27.7%; Pred. No. 2.8e-34;
Matches 208; Conservative 137; Mismatches 321; Indels 86; Gaps 20;

Qy 13 TAATVLAALSSSV-----PAAQTADLETVHIKQSRYSNAIVTEKNGDYSSFAVTVGTKI 66
Db 15 TAAPSLLAMGMAFPAFAAE---DTVIVEGETTADAVNREEQ-DYSMKTAAAGTKM 69
Qy 67 PASLREITPQSVSIITNQOVKDRNVDTFDOLARKTPLGLRVLSNDDGRSSVYARGVEYSEYN 126
Db 70 PMTQDIPQSVSIVSOORMEDQQLQTLGEVMTWTTLGISGQADSRIYSYSRGFEIDNYM 129
Qy 127 IDGLPAQMOS---INGTLPNLFAPDRVEVMRPGSLFDSGEMGGIVNLVRKRP--KAFQ 182
Db 130 VDGIPTYFESRWNLGDALTDTALYERVEVVRGANGMLTGTGNPSASINMIRKUAITSREFK 189
Qy 183 GHAAAGFTHQYKAEDVSGSLNSGSRVGRVMAQTGVAAPRAEKNNRHEFTYAAADW 242
Db 190 GNVSTEYGSWNKQRYMDLQSLPTADGNVRGRIVAGYQNNDSWLDRYNSEKAFPSGIVDA 249
Qy 243 DINPDTVLGAGLYLQQRHL--APYNGLPA-DANNKLPSPQHVFGADW-----NKFK 292
Db 250 DLGTTNLSAGYEQKIDVNSPTWGLPVRWNTDGSKNYSRSTAPDWAYNNKEINKFF 309
Qy 293 MNSHDVFPADLKHYFGNGGYGVK---GMYSRDRDASNYAFAGSKLGMKTP--AGRPOCNT 347
Db 310 VTLKQRFASWQATLNATHTEVEKFDKMMYIDALVDKETGLTVSPYCASYPVVGTCWNS 369
Qy 348 ADDKACAVGLGTE-----IKQALAFDASYSRPFRLCNTANEFVIGADYNRRSTNEQG 401
Db 370 GKRKVDALDLFADGAYELFGRQHNMMFGYSKQNNRYFSAWANVPDDTIGNFSAFNGNF 429
Qy 402 RTLYARGGLALNEFRSIPQVDLIANARKGVGYSHTVATENLDEFGIYKGSTPHPADGL 461
Db 430 PTHWAQNLQADD-----TTHMK--SLYAAATRISLADPL 462
Qy 462 SLIGGRGLGHYKIESGEGKTLHKASKTKFTGYAGAVDLDNNSLYLSLSQLYTPQTNLD 521
Db 463 HLILGARYTNVRVDT----LTYSMEKNHTTYPYAGLIYDINDNWSAVASYTSIFQPNKRD 518
Qy 522 ADGKLLKPROGQFEGVYKGSYMDRLNARVSFYRMKDKNA--APLNPNKKTRYAA 577
Db 519 KAGQYLAPITGNNYEAGLKSMDWMSRLTTLTSLVFRIBQNNVVAQATTPIPGSNGEFAWS 578
Qy 578 LGRKVMESGVEISGAVTPKQIHHAGYSYLHQSQIKTASNRDDGIFLLMPKHSANLWTTY 637
Db 579 TDGTVSGVFEVNGAITDNWQTFGATRY-----VAEDNEGNANPNLPRTSVKLFTRY 633
Qy 638 QV--TPELTITGGGVNAMSGI-----TSSAG---MHAGGYATFDAMAAYRFTPKLQINAD 688
Db 634 RLPAIPELTGVGGVNWQNRVYKDTTTPYGTFRAEQGSYALVDLFTRYQVTKNFSVQGNIN 693
Qy 689 NIFNRHYARVGGANTNIPGSERTWTANLRY 720
Db 694 NLFDKTYTDNIDGSIVYGAPRNV-SLTANYQF 724

RESULT 9
Q9X6A5 PRELIMINARY; PRT; 734 AA.
ID Q9X6A5
AC Q9X6A5;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Ferric alcaligin siderophore receptor.
GN FAUA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

```

RESULT 10

ID	Q8YV33	PRELIMINARY;	PRT;	863 AA.
AC	Q8YV33;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Ferrichrome-iron receptor.			
GN	ALI2148.			
OS	Anabaena sp. (strain PCC 7120).			
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.			
CC	NCHI_TaxID=103690;			
RA	SEQUENCE FROM N.A.			
RP	MEBLINE=21595285; PubMed=11759840;			
RA	Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,			
RA	Matanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kishida Y., Kohara S., Matsunoto M., Matsuno A., Muraki A.,			
RA	Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,			
RA	Yasuda M., Tabata S.;			
RT	"Complete genomic sequence of the filamentous nitrogen-fixing			
RT	Cyanobacterium Anabaena sp. strain PCC 7120."			
RL	DNA Ref. 8:205-213(2001).			
DR	EMBL; AP003588; BAB73847.1; "			
DR	InterPro; IPR002106; AtrRNA_ligaseII.			
DR	InterPro; IPR00531; tonB_box.			
DR	Pfam; PF00593; tonB_box; 1.			
DR	ProSite; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.			
KW	Receptor; Complete proteome.			
SO	SEQUENCE 863 AA; 95445 MW; 40FP9EC04FE462D34 CRC64;			
Query Match	13.6%;	Score 512;	DB 16;	Length 863;
Best Local Similarity	25.0%;	Pred. No. 6,86-22;		
Matches 182;	Conservative 122;	Mismatches 303;	Indels 122;	Gaps 28;

```
Db 682 LSATLAAYQITKSNILTP-DPDPER---AALDYLIQVGEQSRGIELDVAGEILPGWKAI 737
Qy 602 AGSYLHSQIKTANSRDDGIFLL-MPKHSANLWTTTQVTP-----LITGGG-----VNAMS 653
Db 738 ASYAYTNAEV-TEDNIDIPVGNRLVSPKNOASLTTFEFQNSDLKGLGFLGYVGTFRS 796
Qy 654 GITSSAGHAGGYATFDMAAYRTPKLLQINADNLFNRHYARVGGANTFNIPGSERT 713
Db 797 G-DSANGFEIDPDLRTDAAYIYR-RDGFKAGINIRNLFDTDYIKTSDDGRTFLRRGAPFT 854
Qy 714 WTANLRYSF 722
Db 855 IIGSISWEP 863

RESULT 11
Qy 889-267-2.rapt
ID Q8YTW7 PRELIMINARY; PRT; 820 AA.
AC Q8YTW7
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALR2596.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP MEDLINE=21595285; PubMed=11759840;
RX Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003590; BAB74295.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 820 AA; 91035 MW; BBD0B2486468C451 CRC64;

Query Match 12.9%; Score 486.5; DB 16; Length 820;
Best Local Similarity 24.6%; Pred. No. 2e-20;
Matches 185; Conservative 112; Mismatches 353; Indels 103; Gaps 24;

Qy 15 ATVLAALSSVFAAQ-----TADLETVHIKGRS--YNAIVTEKNGDYSSFAVTGKIP 67
Db 126 AGLILAVESTTATQPPATPTSDAPPAPIAQDDPIELVVTGEODRYVPTASTATKTD 185
Qy 68 ASLREIPQSVSIITNQVKDRNVDFDGLARTKTPGLRLVSLNDDGRSSVYARGYSEYNI 127
Db 186 TPVRDIPQSIQVIHQILEDQKTHIQEVLQNVSGVKNQVGGTD---AGYRIRGFDQ 242
Qy 128 DGLPAQMSING-----TLPNLFAPRVEVMRGPGLFSSGEMGIVNLVRKRPYKA 180
Db 243 DG-----NFRNGFNDTFVSLVDTANIDRIEVLKGPASVLFQAPPGGIINVTQPLRT 297
Qy 181 FQCHAAAGFTHQYKARADYVSGSLNSGVS--RGRVMAQTVGASPRAEKNNRHETFA 238
Db 298 FYAAELVGNVAFYRPSDFISGPIITDGLSLRLNVAAYQNSGSF---RDYNFLERVEFA 354
Qy 239 -AADNDINPDTVLGAGLYLQORHLAPYNGLPADANKLPSPHQHVFVGADWNKFGNSHD 297
Db 355 PVITWNIISDRTSLTFDLEVDNDYLPDRGIPISGRPAP-IPISFVGLP-HVYNDSTFR 412
Qy 298 VFADLKHVFGNGGYGKVMRYSRDRDADSNYAPAGSKLGMKTPAGPGCNTADDKACAVGL 357
Db 413 IGYRLEHDFSKDWQRNAPFSFVSGKSSGTAYGGVDL-----IDDQFAPITV 459
Qy 358 G-TEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRRFSTNEQGRITLYARGGLALNEF 416
```

```
Db 460 SRDEFTRDIYTLQTEVVGVKFTGSI VHQPLIGVELAR---NTWKYTSFDVADPILLDIF 515
Qy 417 RSPQVDLLANARKGVGRGSHVTATENLDEFGIYKSTHPADGLSLIGGRLGHY-KIE 475
Db 516 NPNYDELPAFDPESTFSY-----TTRRDITLGIYVQDQITFADNLKLLVGGREFDAPORKE 570
Qy 476 SGBGKTLHKASKTKFTGYAGAVVYDLNDDNSLSLSQLVTPQT--NLDADGKLLKPERQGN 533
Db 571 EGSEFASSESLSAFSPRIGIVYQIOAISLVASYSSQSKPDRFFGSRASNEPFKTRGT 630
Qy 534 QFEVGYKGYMDDRLNARVSFYRMKDKNAAA--PLNPNKTKRYAALGKRVMEGVETEIS 591
Db 631 QYEVGIKAD-ISEKLSATLAAYEITKTNNVTSDDPNLS---VQVGEQSRGIELDIG 685
Qy 592 GAVTPKQWQHAGSYLHSQIKTASNSRDDGIFL-----LMPKHSANLWTTTQVTPELTIG 646
Db 686 GEIVPGWNIISYTY-----TDAITSKDNTPVGNRIDNVPEHAASLWTSYEL----- 733
Qy 647 GGVNAMSGITSSAGMHAGGYATFDMAAYRFTPKL-----KLOINADN 689
Db 734 -----QSGDLKGLGFLGLYVGDYADVENTSLLSYFRTDSAIYYKRDNRWRLALNFRN 788
Qy 690 IFNRHYARVGGANTFNIPGSERTWTANLRYSF 722
Db 789 LFNETYETTSQARNTI-YPGAPFTVIGSFSIQF 820

RESULT 12
Qy 889-267-2.rapt
ID Q8YXV8 PRELIMINARY; PRT; 885 AA.
AC Q8YXV8
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome iron receptor.
GN ALL1101.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003584; BAB73058.1; -
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 885 AA; 98141 MW; AF7F6BAE0A37FE08 CRC64;

Query Match 12.9%; Score 486.5; DB 16; Length 885;
Best Local Similarity 23.9%; Pred. No. 2.2e-20;
Matches 179; Conservative 117; Mismatches 343; Indels 109; Gaps 26;

Qy 7 VFRINMTAATVLAALSSVFAAQADLETVHIKGRSYNAIVTEKNGDYSSFAVTGTKI 66
Db 191 IFEVAAPTAPTAPQESQVQPEQPS-----ARSEEPIELVVTGEQDGYKVEDATGGT 244
Qy 67 PASLREIPQSVSIITNQVKDRNVDFDGLARTKTPGLRLVSLNDDGRSSVYARGYSEYNI 126
Db 245 NTPIRDTPFSIQVPEEVIKDDQVQRVTDALRAVPG--VISQDAPVSA-----FESFN 295
Qy 127 IDGLPAQMSINGTLPNLFA-----FDRVEVMRGPGLFSSGEMGIVNLVRKRP 178
```

```

Db 296 FRGSSSFLRNGRIDETIGTAGSGVANIERIEVLKRGAGALFSGSGRGTVNITVKOPL 355
Qy 179 KAFQGHAAAGFTHQKQYKAEADVSGSLNSDGSVRGVAQAQTVGASPREAEK-----N 230
Db 356 SSPFEIEGTGVSFDIYEGRVDLTGPILRND-----TLVYLTAS---ASLGSFIDFVD 408
Qy 231 NRHETFFAADM--DINPDYVLGAGYLYQQRHLAPYNGLPADANNKPLSLPQHVPVQADW 288
Db 409 NERYFISPVLTWLDKXTNLTLEAEYLSTKN--PNYNGLPF--LGTVLPNPEGEIPLSNTL 465
Qy 289 NK--FKNSHVDFA---DLKHYFGNGGYGKVMYSRSDADSNYAFAGSKLGMTPRGRP 343
Db 466 NEBFDKNDROVYRLGYNFHRFSENNQFNSLPAFOEYQNEAIPLLELL-----516
Qy 344 GCNTADKACAVGLGTEIK--QKALAPDASYSRPRFLGNTANEFEVIGADYRFRSTEGQ 401
Db 517 -----EDNRITLVRAQSLSTRNYYVLTNTNVGDFPKGSIHKLPLFGDLLRDTRRDT 571
Qy 402 RTTYIYAGGLALNEFRSIPQVDLIAMARKVGVSHVATNTLDEFGIYGKSTHPADGL 461
Db 572 QDPAIE-----PINLFNPYGTQTIGSSAPSSNFTSNT-----DGLGIYLDQMTLADNF 621
Qy 462 SLIGGGRGLG--HYKIESGEGTLHKASKTKFTGYAGAVYDLDNNSLYLSLSQLYTPQTN 519
Db 622 KILGGFRDIIISQKQEDNDGETINFLODEAPSPRIGLYOPTKNI SLGYSYSRSFTQYVG 681
Qy 520 LDADGKLLKPRQGNQFEVGYKSYMDRLNARVSFYRKQKNAAP--PLNPNKKTYYAA 577
Db 682 TSPDQRLFOBERGQYEVGIKGMDLDRKLTSTLAFYQITRNNVSNBPDNG----FLIQ 737
Qy 578 LGRKMEGVETESGAVTPKQIHAGYSYLSHQIKTASNSRDGIFL--LMPKHSANLMTT 636
Db 738 TGEORSOGIEIDILAGEILPGKIIIGYAY--SDAKTQDQFEGNLLNNTKHAFLMTT 795
Qy 637 YQVTELTIGGVNANGITSAGMHAG-----YATPDMAAYFTEPKLX 682
Db 796 Y---ELQSGN---LQGLGFLGLYIGERGDLNSFELPSYFRDAMFYR--RUNFR 846
Qy 683 LQINADNIFNRHYARVGANTFNIPGS 710
Db 847 AALNINLFLTEYETAFDALSV--VPGA 873

RESULT 13
Q8YUK0 PRELIMINARY; PRT; 854 AA.
AC 08YUK0:
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, last annotation update)
DE Ferichrome-iron receptor.
GN ALR2211.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; Pubmed=11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AP003588; BAB7310.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 854 AA; 94535 MW; A379CEFC075552C7 CRC64;
Query Match 12.8%; Score 483.5; DB 16; Length 854;

```

```

Best Local Similarity 24.1%; Pred. No. 3.2e-20;
Matches 183; Conservative 134; Mismatches 328; Indels 113; Gaps 29;

Qy 18 LAALSSVFAQAQTADLEFVH-----IKGQRYNALVTEKNGDYSPAVAVGKRIASL 70
Db 157 LAPASTIPAEQPEQPTSETPPAQPTAENNEPELAVTGNQDQRYVEESSTATIKIDPL 216
Qy 71 REIPQVSIITNQCKDRNVDPDLAKRTPLGLVLSNDGRSS--YVARGYSEYNIID 128
Db 217 RDI PASVQVLPKEIIDQRYVRLNELANNVSGVQGGYGGSLSSGFIQF---ESGFE 273
Qy 129 GLPAQOMOSINGTLPLNLA--FDREYVMRGP--SGLPDSSGEMGIVNLYRKPTKAFQGHAA 186
Db 274 GLRNGFDFGFTSPRDANIERVEFLGPASVLYGSANNCGVYNTITKXPLPDPYSRVG 333
Qy 187 AGFETHQYKAEADVSGSLNSDGSVRGV--MAQTVGASPREAKNNHETFYAA--ADWD 243
Db 334 MTIGSYPTFRPTIDFTPLDDSKSVLYRLNVAENSGSPFDFTEN---ESFFISPVTVN 390
Qy 244 INPDYVLAGAGLYQQRHLAPYNGLPADANNKPLSLPQHVPVQ--ADWKKFKNSHVDVADL 302
Db 391 ISPTSMTFEYEVQKNVYTFDRGLL--PGNTFPOIPISRLGEGFRNNAEISVFTYNL 448
Qy 303 KHYFGNG----GYKVMYRSDRDADSNYAFAGSKL--GMKTPAGRPGCNTADKACA 354
Db 449 EHQPSDMKFRQGFNVTSIGNTRIA--RNTNFSPEFLDEDDQTLPR---TSETSDEQEN 504
Qy 355 VGLGTEIKQALAPDASYSRPRFLGNTANEFEVIGADYRFRSTNEQRTLYIYAGGLALN 414
Db 505 ISLQTEVSGK-----FNTGSI RHNVLLGVELAKKFT-----Y 537
Qy 415 EFRSIP--QVDL--IANARKGV--RGVSHVATENLDEFGIYKSTFHPADGLSLGG 467
Db 538 DFFSAPIASIDIRNPVYGAQPTFDRSFAGEYGDNL---AVYQNLIEFTPNKLALAG 594
Qy 468 R---LGHYKIESGEGTLHKASKTKFTGYAGAVYDLDNNSLYLSLSQLYTPQ--TNLAD 523
Db 595 RFWIDSFDRDPVSNVTNNEVSESNFSPRVGIYQPTNSYSLVAMTNSFNPQFGRSRT 654
Qy 524 GKLLKPRQGNQFEVGYKSYMDRLNARVSFYRKQKNAAP--PLNPNKKTYYAALGRVM 583
Db 655 GESFKPETSQFEVGIKQEFDFDKLSATLAFYDITKNNVLTTPDNDNFS--VOIGQKS 712
Qy 584 EGVETEISGAVTPKQIHAGYSYLSHQIKTAS-----NSRDDGIFLMPKHSANLMTTQ 638
Db 713 RGLHLDIAGELPGKIIAYIYIDSVSXNDLERLNDLSGV---PNSASLMTTYE 768
Qy 639 VTEPLTIGGVNANGITSAGM-----HAGGYATPDMAAYFTEPKLKIQ 684
Db 769 FQK-----GSLAGLGFGLVYDEREATLPTIKIIPSYRTDASIFYR--RDNMRAA 819
Qy 685 INADNIFNRHYARVGANTFNIPGSERTWTANLRYSF 722
Db 820 INIKNLPDTEYE---SQSEYLVPAAPLTVIGTISFEF 854

RESULT 14
Q8YMK9 PRELIMINARY; PRT; 858 AA.
AC 08YMK9:
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, last annotation update)
DE Ferichrome iron receptor.
GN ALI4924.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; Pubmed=11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,

```

RA	Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,	
RA	Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,	
RA	Yasuda M., Tabata S.;	
RT	"Complete genomic sequence of the filamentous nitrogen-fixing	
RT	cyanobacterium <i>Anabaena</i> sp. strain PCC 7120.";	
RL	DNA Res. 8:205-213(2001).	
DR	ENBL; AP003598; BAB76623.1; -.	
DR	InterPro: IPR000531; TonB_boxC.	
DR	Pfam; PF00593; TonB_boxC_1.	
DR	Receptor; Complete proteome.	
KW	SEQUENCE 858 AA; 94913 MW;	8F227AFCEA3E7818 CRC64;

Query Match	12.7%;	Score 481;	DB 16;	Length 858;
Best Local Similarity	25.2%;	Pred. No. 4.5e-20;		
Matches 189:	Conservative	127;	Mismatches 339;	Indels 96;
				Gaps 27;

Qy	17	VLAASSVFAAQADTLETHIIKQ-----R5NAIVTEKNGDYS5FAVTVGTKIPASL	70
Db	159	LIVAT5AKVAMOPP--ETPAIQEPAAPQDEPIELWTTGEQDQYRVPAA5VGTETDTEL	216
Qy	71	REIPOSVSIITNOQVDRKNVDTFDQLARKTPGLR-VLSNDDGRSSVYARGYEVSEYNI-	127
Db	217	RDIPQ5TOVVRQVLQEQRATRGLDALRNVGSVNPTRGSGDRADSTIIRGFELFSGNVLN	276
Qy	128	DGLPAOMQ5INGTLP---NLFAFDRVEMRGP5GLFDSGEMGGIYNLVRKPTKAFQGH	184
Db	277	NGLP-----DRTLITETRDLYNVERVEVLKGPASVLYGLGNPGGTVNIVTQPLANPPYN	330
Qy	185	AAAGFTHQYKABADVSGSINDSGSVGRVMAQTVCASPRPAEKNRHHETFYAAADWDI	244
Db	331	IEATVGNVNDLYRGDIFSGPLNDSKTYLLRLNLAYQNSGSYIDFVGNRSFFIAPVISAAL	390
Qy	245	NPDTVLGAGVLYQORHLAPYN--GLPADANKLPS-----LPOHVFVGADWNKFKMNSH	297
Db	391	GKNTLTLTPEGEYSOKTIDSRTVVVLPA-VGTVLPGDGRRIIPNRVTYEPGEDTQIETTR	449
Qy	298	VFADLKHVFGNGGYKVKGMRY5-DRDADSNYAFAGSKLGM--KTPAGRPGCNATDDKAC	353
Db	450	LGYRLEHRF5ENWSLRNDFRVTPEHNNADNQAFP--LGLDADNRNTANRSTYSSES	505
Qy	354	AVGLGTETKOKALAFDASY5RPPRLGNWTANEFVIGADYNFRSTNEQRTTYARGGLAL	413
Db	506	IYNLTJTDI-----SGRFFTG5IEHQLLFGVNM5RFDNF-----NFGIDL	545
Qy	414	NEPSRIPOVDLIANARKGVGYSHTVATEN--LDEFGIYK5TFHFPADGLSLIGGRLG	470
Db	546	AELTPLDIYNPV--YRQPIVGRIDAVYEDNSRLTDTLGIYIQOIKFAENFQLLGERFD	603
Qy	471	HY-----KIBSGEGKTLHKASKTKFTGYAGAVYDLNDDNSLYLSLSQLYTPQT	518
Db	604	LFTQKNKNFLNDTTEIQAGDA-----FTRPVGIYVKPISAI5LYASYSQSFPNTE	653
Qy	519	NLDADGKLLKPRQNGQFEVGYK5YMDRLNARV5FVRMKDKNAAPLNPNNKKT5YA-A	577
Db	654	GRSADGNLFPERGTOVEVGVKAD-LNDRISSTLSLRT5R5NLLT--TDPNN--SRFSIQ	709
Qy	578	LGRVMEGVETEISGAVTPKWQIHAGYSYLHSQIKTASNRDDGIFLLMPKHSANLWTTY	637
Db	710	TGEORSQIEFDIAEILPGWKIFAGAYTDAEIVEDNTVTSNGRLTNVPEHFN1WTTY	769
Qy	638	QVTPELTIGGGVN-AMSGITTSAG-----MHAGYATFDAMAAYRFTPKLKLGINADNIF	691
Db	770	EISSGDPRGLGFLGLFVIGDRAGDLONSFEVPSYLR5TASIFVR-RDLR5PALNMKNIF	828
Qy	692	NRHYARVGGANTFNI5G5ERTWTANLRY5F	722
Db	829	DTDFV5V5TNRD-FVLRGDPPTI5G5T5WDF	858

RESULT 15	
Q9AC38	/
ID Q9AC38	PRELIMINARY;
AC Q9AC38;	PRT; 851 AA.

DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	TonB-dependent receptor.
GN	CC0028.
OS	Caulobacter crescentus.
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC	Caulobacter.
OX	NCBI_TaxID=155892;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 19089 / CB15;
RX	MEDLINE=21173698; PubMed=11259647;
RA	Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Hatt D.H.,
RA	Keloney J.F., Smit J., Craven M.B., Khouri H., Shetty J., Barry K.,
RA	Utterback T., Tran K., Wolf A., Vamachavan J., Ermolaeva M., White O.,
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT	"Complete genome sequence of Caulobacter crescentus.";
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR	EMBL; AE005677; AAK2016.1; -.
DR	TIGR; CC0028. -.
DR	InterPro; IPR000531; TonB_boxC.
DR	Pfam; PF00593; TonB_boxC; 1.
KW	Receptor; Complete proteome.
SO	SEQUENCE 851 AA; 90124 MW; B75B166237F5CD92 CRC64;

```

Query Match      12.7%: Score 479; DB 16; Length 851;
Best Local Similarity 24.7%: Pred. No. 5.8e-20;
Matches 139; Conservative 126; Mismatches 34; Indels 138; Gaps

Qy 10 INMTAAVTVLAALSSVFAAQTADLETVHHKGRSSYNAIVTEKNGYSSP-AVTV----- 62
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 95 LSLAAVAGVAGL-----LAQAA-----VAGPDSADAAVTTANADRSVSSVTTDAKRA 144

Qy 63 ---GTKIPASLREIPGSVSIITNQVKDRNVDTFDQARKTPGLRVLSNDDGR-----SSVY 116

Db 145 DPSSSKPTA PLVDTPKSVTVIPAKIIETQATAATSLADILTSIGITFGAGEGQPLADRPFF 204

```

Qy	117	ARGVEYSEYN--IDGLPAQMQSINGTLPNLFADPDRVENVMRGSGULPDSSGEMGGIYNLVR	174
Dd	205	IRG-QASANNVFVDG---VRDSCGVQRIFENLEQVEVVYKGPDSAYGGRSGGGGSINLSS	259
Qy	175	KRPTKAFQGHAAGFTHQYKAEDAVDSGLNSDGSRVGRVMAQTVGASP-RPAEKNNRH	233
Dd	260	KSPKADSFARSGVGVDYAVRATADLNHALNESVAURLNLIA-TQGDTPGRKSVSFDR-	317
Qy	234	ETFEYAADDWINP-----DTVLGAGLYIQORHLAPYNGLPPADANNKL P-----SL	278
Dd	318	-----WGVPFLAIGLDGDTQLTASYVHLEGDDTPDYGVPLLTKTTPRTASGILDV	369
Qy	279	PQHVFUGADWNKFPMNSHDV--FADLKHFPGNGGYCKGVMRY S-----DRDADSNVA	328
Dd	370	RRRSFYGVASRDYQKTKSDIATEFA-IDHRIDEITLNRQVRYRYSKSLNDYIVTNPGDGGAA	428
Qy	329	-FAGSKLGMTKPAGRPGCNTADDKACAVGLGT E I --KOKALAFDASYSPRFFRLGNATANEF	385
Dd	429	QFVOGQWMMXR-----GTKRNPNPTETVAADVTLDHGKKTFLGLEHSFDVGLELSREEN--	481
Qy	386	VIGADYNRFRSTNEQ-----GRTTLYARGGLUALNEFRSIPOVDL-----IANARKGVRCY	435
Dd	482	-LNATYSTFTTSGAACPFGTIIATLASLAGDCDCTLVYKPKMDKAWTGVIINRAPAARNV	540
Qy	436	SHTVATENLDFEGYKGSTPHPADGLSLIGGRGLGHYKYESGEKTLH-----	483
Dd	541	AKTTA-----LYGPDTVKFGEKVLLNLIJLRHDIRESKGVDDVATTQANGVFTSVTYTP	592
Qy	484	KASRTKETGY-AGAVYDLNDNNNSLYLSLOLYTP-----QTNLDADGK-----LLKP	529
Dd	593	RSGSWAEFTNYOVGLVYKPTPGSSLXYVSYSTASTPFGISAGDONISNTATGTGNLATVOLSP	652

```

QY 530 RQGNQFVGVKSGSYMDRLNARVSFPRMCKNAAPLNNPKTRRYAALGKRVMEVETE 569
DB 653 EDSESEFAAKAFVFDHTLALSNALPQTSRKNAQIQID----AFTYAQVGEVVEKFEFG 708
QY 590 ISGAVPKMQOIHAGSYLHSQ-1KTASNSRDCIFLL-MPKHSANLMTTYQVTPPELLTIG 647
DB 709 VSGNITPKMQVREGIYYMSELYKGYTSVQGDPLANTPKHSISSFTYKVRKIALG 768
QY 648 GV----NMSGITSSAG-----MHAGYATPDMAAAYRTPKILQINADNIPNRHYAR 698
DB 769 GAVHVSXSGGNGGAGGASRIYAPAYRKYDAPASMAVSTGVDLOLNTLIDERYIAR 828
QY 699 VGCANTFN-IPGSERTWTANLRY 720
DB 829 TNGVHADPAFGROAILTNVXY 851

RESULT 16
QYV06 PRELIMINARY; PRT; 863 AA.
AC Q8YV06;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Ferrichrome iron receptor.
GN ALR2175.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxId=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriuchii M., Ishikawa A., Kawashima K., Kimura T.,
RA Kiehlida Y., Kohara S., Matsumoto M., Matsuno A., Muraki A.,
RA Kazakazi N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003588; BAB73874.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 863 AA; 95343 MW; 3EDA3D64447D2372 CRC64;

Query Match 12.6%; Score 475; DB 16; Length 863;
Best Local Similarity 25.5%; Pred. No. 1e-19;
Matches 176; Conservative 120; Mismatches 301; Indels 94; Gaps 26;

QY 47 IYTKNGDVSSFAVTYGTIPASLREIPQSVSITNQVCKDNVDFDQARKTG-LRV 105
DB 201 VTTGADSDSRVNNASATRTDPLNDIPQSVQVPOQVLDQOVNLNALQNVSGVIT 260
QY 106 LSNDDGRSSVARGYEV-----SEVNIIDLPAOMOSINGTLFNLFAFDREVMRGPGLF 160
DB 261 ASNYQSFASFTIRGFNSFDQGNFTNGLGYRFGS-QGT-NFSNIERILEVLRGPGSVL 317
QY 161 DSSGEMGCVLVNKRKPTAFPGHAAAGCTHKQYKAEDVSGSLNSDGSVGRVMAQTV 220
DB 318 FGSGNPGGCTINIVTKQPSLEPFYSVEAAGSYDFYRGALDLSGPIDDSGPYTALYR-----L 372
QY 221 GASPRPAEK--NNRHEFYAA--DWDINPTVAGAGLYOQRHAPVNGLPADANNLT 275
DB 373 NASYEKADNPFVDFNRENFSVASTLSFALGENTITLTLDOENKVNQGYTVGPA-VGTVL 431
QY 276 PSL-----PQHVFGADWNKFKKNSHDVADLKHFGNGGYGVKVRYS---DRDASNYA 328
DB 432 PVLNRIPIRNRISIGQADSTYSPEIVRVGYNLHKFSSEDLNNAFYSHFYKTRDTYFA 491
QY 329 FAGSLGKMTKPRGCGNADDAKACAVGLGTEIKQALAFDASYSRPFRLGNTANEFYIG 388
DB 492 NA--LDPDQRTLQRGVQDADRDYQTYDLSITNVGK-----FSTGSIKHQILFG 537

```

```

QY 389 ADVNRFRSTNEQRTTLVARGGLANEFPSIPOVD-----LIANARKGVGSHVTATENL 444
DB 538 VLSRY-----DQVSEYERBTGTPLDLFPNDRPERFEVILFA-----GNEATLTL---- 582
QY 445 DEFGYKSTFHPADGLSLIGGRGLGHYKIESGE-GKTLHKASKTKFTGYAVAYDLND 502
DB 583 DSLGVYIQQVTFAEENFKLLGGRFDIFQYTKTRDLSNTEQOSGSAFSPRVGIYVQPI 642
QY 503 NNSLYTSLISQLYTPQTNLDADGKLKPRQGNQPEVGYKSYMDRLNARVSFPRMCKNA 562
DB 643 PLSLVAYSRSFTPTIGRASDGEQPEPGRTGYEIGVKN-DINEXLSATLAFYDLTRSNV 701
QY 563 AA--PLNPNKTKRYAALGKRVMEVETISGAVPKMQOIHAGSYLHSQIKTASNSRD 620
DB 702 TINDPANPEFS-----IQTEQNSRGIELNVAGEILPGNNIINGIATYTDARI-TQDSLPI 756
QY 621 GIEFL-LMPRHSANLMTTYQVTPPELLTIGGAVNMSGITSSAGMHAGG----- 665
DB 757 GNRLLNVPEHSLSLWTTY-----ELQKN-----LQGLGFGGLGFYIGDRQGLANSFNLS 808
QY 666 YATPDMAAAYRTPKILQINADNIPNRHY 696
DB 809 YLRDPAALFYK-RDRFRALNIRNLDLKYF 838

RESULT 17
QYV06 PRELIMINARY; PRT; 733 AA.
AC Q926C7;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Putative ferrichrome-iron receptor precursor protein.
GN PHU01 OR R02223 OR SMC01611.
OS Bacterium melliloti (Sinorhizobium melliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxId=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Leleau V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purrelle B., Rameberger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium melliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591790; CAC46802.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 733 AA; 79611 MW; 36D5C049C9B3786E CRC64;

Query Match 12.5%; Score 471.5; DB 16; Length 733;
Best Local Similarity 24.6%; Pred. No. 1.3e-19;
Matches 192; Conservative 114; Mismatches 338; Indels 135; Gaps 30;

QY 12 MTAATVLAALSSVFA-----AQTADLETVHKKQGSNAIYTEKNGDVSSFAVTYGTIP 67
DB 14 LASGVALLPPLMMSGIALAQEGNATQLERIIVEGGNAGASATGVPDGVAAATATGSKTA 73
QY 68 ASLREIPQSVSITNQVCKDN-VDTFDQARKTPGLRVL-----NDGRSSVARGYEV 122
DB 74 MPLNEIPQSVSVGHELDDBRAVNVKVDALRYTPG-VLSAPCTPDDTDMFYRGRDA 131
QY 123 SEYN--ITGLPAQOSINGTLFNLFAFDREVMRGPGLFDDSGEMGCVLVNKRKPTXA 180
DB 132 AQTGFLFDGLPLFSFGFNQVDFPMLERVELVKGPAVSLVYGSGNPGGIIINLISKRPIDE 191

```

181 FQHAAGFGTHKQYKAEADVSGLSNDSGVRGMAQTVGASPRPAEKNNRHETTYAAA 240
192 PLYYTEVINGNSNGNAFTGFDNKLDDGTVRVRLTKVAGGDN-----YSDY 239
241 DWDINPDTVLGAGLYQOORHLAPYNGLPADANNKLSLPQ-HV-----FVG-----ADWN 289
240 SEDLR-----GFILPQVTVAPDDATSLTVFGLQSLDQVHVGVNGFLPYVGVTEADPFG 292
290 KFKMNSHVDVADL-----KHVFGNGYKVKVMRYSDRSDADSNYAFAGSKLG 335
293 KIDRDVAYSEPDIDEGSYTOQMLGYEFKDFDNGWTTQNNARVANLHKHKEPYTYGVYG 352
336 MKTPAGPGCNTADKAC-AGVLGTEIKQALAPDAGSYSPFRFLGNTANEFVIGADYNRF 394
353 -----GAPTPGPDYLLNRIGFPAATSKVDSFSDNRTETDFDGLATTHFLAGLDYKY 404
395 RSTNEQ---GRTTLYARGGLALNEFRSIPQ-VDLIANARKGVRYGSHVTATENLDERGIY 450
405 RLDHIQACCGATPI-----SATNPVYGTPOGANFV-----YLDQIVTQ--QQIGLY 448
451 GKSTFHPADGSLGGLGHLGHYKIESGEG-KTLHKASKTKETGAGAVYDLDNNSLYLS 509
449 AQDQIRFGDGLVTLNGRYDYVDTKSDAAIGTSYESNDGAFSGRAGLAIEFDNGLTPYVS 508
510 LSQLYTP-----QTNLDADGKLLKPRQGNQFVG--YKGSYMDRLNARVSPYR 556
509 AATFPNPLVGTGSDPSDPTN-PAKNVALEPEEGYQYAGVKYEPESFDGLLTASVFQIT 567
557 MKDKNAAP---LNPNNKKTRYAALGKRVMEQVETEISGAVTPKQIHHAGSYLHSQIKT 613
568 KQNVSIAPGPFVN-----SOLGEVRSRGEVLEGKINLTNNKIIISAFSYTDLEYTE 619
614 ASNSRDDG-IFLLMPKHSANLWTTQVTPELTIGGVNAMSGITSSAGMHAGGAVTADAM 672
620 DLNASLIGNTPVLIPETQASLMDY-----TVANG--TFEGVSLGAGVRYOGESWADAE 671
673 -----AAVFTPK-LKLIQINADNINRHHYARVGGANTFNPISGERTWTANL 718
672 NTKKVPAATLVDAATRYEKNDWTASLNVANLFDKEVAGCQGLQICGY-GESRTFTLKL 729

RESULT 18
Q8YTP0 PRELIMINARY; PRT; 867 AA.
AC Q8YTP0;
DT 01-WAR-2002 (TReMBLrel. 20, Created)
DT 01-WAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALL2674.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003590; BAB74373.1; -
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; -
KW Receptor; Complete proteome.
SQ SEQUENCE 867 AA; 95468 MW; 5D21556D8E781FPA CRC64;

Query Match 12.4%; Score 467.5; DB 16; Length 867;
Best Local Similarity 24.7%; Pred. No. 2.8e-19;
Matches 186; Conservative 127; Mismatches 264; Indels 175; Gaps 38;

61 TVTGTIPASLREIPQSVSIITNQOVKDRNVDTFDQARKTIPGLRVLSNDDGRSSVYA--- 117
201 SVGTTRDTPLIDVPAIQIPOEVIKDQTRSLGEVLKNTS-----SASSGRTSQAPAL 255
118 ---RGYEYSEYNIDGLPAQMOSINGTILPNLFAFVRVEMRGPGLFDSSGEMGGINLV 173
256 TPVIRGFESRNLRLNGLRDSDSLRFQSEIANV---ERVELKGPASVLFQGGDLGGVNLV 312
174 KRKP--TKAQF-GHAAAGFTHKQYKAEADVSGLSNDSGVRGMAQTVGASPRPAEK- 229
313 TKOPLNTPYISGYGVQGFGLH---RPTIDFGPLDKOG-----VAYRLNAAQTAESF 363
230 ---NNRHETTYAADWDI--NPDTVLGAG--YL-YOQRHLAPYNGLP----- 269
364 KDPENSESFPIAPVVKLIGNENTNLTAETYLKYSFETAP--DLPASGTVISNPNGRVS 421
270 -DANNKLPSPQHVFGVAGDWKFKMNSHVDVADLKHVFGNGYKVKVMRYSDR---DADS 325
422 RETNLGEPSE-----SESLVTRL-----GY-QLDHLRLNDNWTIKSEF 459
326 NYAPAGSKLGMKTPAGR---PGCNTADDKACAVGL---GTEIKQ-----KALAFDA 370
460 STAF-----LDVPESTVVLPIANSSTN---GLNRDGRTLRRLVENSPSTSLTFNN 509
371 SYSRPRLGNTANEFVIGADYNFRSTNEQGRITLYARGGLALNEFRSIPQVDLI--ANA 428
510 SLGKFKTGSIEHSLFGVEV---TRETEDEL-----DFQLRDIIDIFNPVYR 555
429 RKGVRGYSHTVATENLDE--FGIYKSTFHPADGLSLIGGRLG---HYKISGEGKTL 482
556 PESVSSFAIPFGNTNTEKNSIGIYAQOISLSKNIILVLGRLDFVNVQDYEDLLSEESF 615
483 HKASKTKFTGYAGAVYDLDNNSLYLSLSQLYTP-----QTNLDADGKLLKPRQG 532
616 ER-NDTVFSPRVGIVYKPSNLSIYASYSRSTFPVQGRVLTNTGTTTGVGEPFEPERG 674
533 NQFEGYKGSYMDRLNARVSYFVRMKDKNAAA-----PLNPNKKTRYAALGKRVMEQVE 587
675 TQYEVGLKANLGLRSLTTLAFYNLERTNVAAQGLSEPLS-----QIQIGKQRSQIG 727
588 TEISGAVTPKQIHHAGSYLHSQIKTASNSR-DDGIIFLLMPKHSANLWTTQVTPELTIG 646
728 LDVAGEILPGNHLTASAYTDSKITEDSRPEFQDQQLQNVPRNSFGLWSTY----ELQAG 783
647 GGVNAMSGITSSAGMHAGG-----YATPDMAAAYRTPKLIQINADNIFN 692
784 ---SLKGLGFLGVFTQGERQGLRLNTFTLPSYLRTDASIFYR-RDKFRAAINQNLFD 838
693 RHYIARVCGANTF--NIPGSERTWTANLRYSF 722
839 ENYYE---GARDIVRVIFGAPFTLTGVSVEFF 867

RESULT 19
P72602 PRELIMINARY; PRT; 828 AA.
AC P72602;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Ferrichrome-iron receptor.
GN FHUA OR SLL1406.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasaki K., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpoo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90899; BA16602.1;
 DR InterPro; IPR00531; TONB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR Complete proteome.
 SQ SEQUENCE 828 AA; 91420 MW; 315D8491FA82FE3 CRC64;

Query Match 12.3%; Score 465; DB 16; Length 828;
 Best Local Similarity 23.1%; Pred. No. 3.7e-19;
 Matches 182; Conservative 119; Mismatches 303; Indels 184; Gaps 30;

```

QY 17 VLAALSSSVFAA---QTADLETVHIKQGRSYNAIVTEKNGDVSSF--ATVGTGKIPASLR 71
DB 143 VLSLTSSSTALAPENPESETEV-----ATQEGGSEASYFVSPSASTATGLDTPLL 192
QY 72 EIPQSVSIITNOQVDRNVDFDQARKTGRLVLSNDGR--SSVYARGYSEYNI--- 127
DB 193 DIPQIQVVPQGVLDNRVTELGRLAQTVPG---VSPAGRGTSVFGPGLIRGFPVNN 249
QY 128 ---DGLPAQ-MOSTINGTLPNLFAPRVEYMRGPGSLFDSGEMGIVLVKRRPTKARQG 183
DB 250 IFRDIPYQSLAPLNTT-----DIEQIEVLKGPSSIVFGAGEPGGSIMLSKKPLDEPY 304
QY 184 HAAAGFTHKQVKARADVSG-----SLNSDSVNGRVAAQTVGASPRBA 227
DB 305 NNAVSLGYNDRDLVDLSGRLPRAIDTVNRLNVSEYTSISFDFYGDLMVVSF--- 361
QY 228 EKNNHETFFVAAADMDINDPVLGAGYLYOQRHLAPYNGLLPADANNKLPSPLOHFVNGAD 287
DB 362 -----TLTNIGBDTKLNTYGGYTTNRRLDSEIPAR---NLADEPSNFFLEGR 407
QY 288 WNKFMNSHDVFPADLKHYFGNGGYGKVRYSDRDADSNYAFASKLGMKTPAGPGCNT 347
DB 408 FSKPEODQYLLIGYTEFNHPFNENLKLRLHAMQY-----LAVAPRYA-----PLRDF 452
QY 348 ADDKACAVGLGEIKQKALAPDASRPRRLGNTANEF-----VIGADYNFRSTN 398
DB 453 FDEDT-----GELNRFYYGGGNTQRFETNAELIGEFTYGPVKHRLVFGLEYRNDTEP 506
QY 399 EOGRTTLVARGGLALNEFRSIPQVDLIANARKGVGSHT---VATE-----NLDEFGY 450
DB 507 EFGFSNTFA---PINVENPV-----YNTPEPIAPPEFRDDQVNRFAVY 547
QY 451 GKSTFHPADGLSLIGGRLGHYKLESSEGGKTLAKASKT-----KFTGYAGAV 497
DB 548 LQDQMDLFDNLKLVGLHY-----DSATQNRSTQSIITDREEFNQTDNLQTRVGII 599
QY 498 YDLNDNNSLYLSLSQLYTPQ--TNLDADGKLKPRQGNQFEGYGSYMDRLNARVFFY 555
DB 600 YQPIPTVLSYTSYTSFNPSPASISNADGSTFDPQTGRQFEGVAVAD--ITTKLSTFSAF 658
QY 556 RMKDKNAAAPAPNPNKTRRYAALGKRVMEGYETEISGAVTPKQOIHAGYSYLSHOKTAS 615
DB 659 DIRQGNVET-IDPANLLETIQT-GEQTSRGVELLYGGEILFGMNIVTGISYLDARVQDN 716
QY 616 NSRDGIFLMPKHSANLMTTYQVTPPELLIGGVNAMSGITSSAGMAG----- 664
DB 717 TDIVDNTLSNVPNSQFSMTIYEI-----QSGNIGQSGFGGLFVVDQREGDL 764
QY 665 -----GYATFDMAAARFTPKLQIINADNIFNRHYAARGANTNI---PQSEPTW 714
DB 765 DNTFVLPSYFPTDAIFVR-BENNELQINIENTYTOYLA--SNDPLSVYGPAPPTV 820
QY 715 TANLRYSF 722
DB 821 VGKIGVTF 828

```

```

Q8XR06
ID Q8XR06 PRELIMINARY; PRT; 695 AA.
AC Q8XR06;
DT 01-MAR-2002 (TRENBLREL 20, Created)
DT 01-MAR-2002 (TRENBLREL 20, Last sequence update)
DT 01-JUN-2002 (TRENBLREL 21, Last annotation update)
DE Probable ferrisiderophore receptor protein.
GN RSP1062 OR RS02469.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Geuzy J., Mangelot S.,
RA Ariat M., Billault A., Broctier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaopin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguler P., Thebaute P., Whalen M., Winkler P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
DR EMBL; AL646082; CAD18213.1; -
DR InterPro; IPR00531; TONB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR Plasmid; Complete proteome.
SQ SEQUENCE 695 AA; 75590 MW; FD7CA6956A201EBE CRC64;

Query Match 12.3%; Score 463; DB 16; Length 695;
Best Local Similarity 24.5%; Pred. No. 3.8e-19;
Matches 185; Conservative 112; Mismatches 336; Indels 122; Gaps 30;

QY 22 SSSVFAAQT-----DLETVHIKQGRSYNAIVTEKNGDVSSFAVGTGKIPASLR 72
DB 5 AQAAPAEBAARAGPTVEAGTLPATVIGQTRSRTS-----YTPVATTKIPAPLRE 56
QY 73 IPOSVSIITNOQVDRNVDFDQARKTGRLVLSNDGRSSVYARGY- YSEYNIDGLP 131
DB 57 IPOSVNVVPRAVIODGALSINDTLRNVPVGSASLGAQRQVITRGSAINQDVQGLR 116
QY 132 AQMSINOTLPNLFAPRVEYMRGPGSLFDSGEMGIVLVKRRPTKARPGHAAAGGT 191
DB 117 DDALYFR---DLSNVERIEVLKGPAAVLVGRGSSGGIINRVTKKPLATPMAEYVLLGT 172
QY 192 HKQYKAEADVSGSLNSDGSVNGRVAAQTVGASPRPAEKN---RHETFYAAADMDINDP 247
DB 173 EGQGRABEDLNTSINDA-----VRALITGAVEDSGFRNDYFLRQALISFPLFNISPD 227
QY 248 T-VLAGGYLYQQR---HLAPYNGLPADANNKLPSPLOHFVAGADWNKFMNSHDVAFAD 301
DB 228 TKLTLQFDYLDHTRIDQGVPSYGRPVN---VPIETRYGSANAAGQANNVETTKSVTGT 283
QY 302 LKHFGN-GGIGKGYKMSYSDRDADSNYAFASKLGKMTTPACRPGCNTAADKACAVGIGTE 360
DB 284 LDRHINDQWSESHSVNRYEYALGNNNYTVVSRYTGAVPVTVLGVNORNSDRGTWQNE 343
QY 361 IKORALAFDASYSRPF--RICNTANEFVIGADYNRFRSTNBQGRITLYARGGLALNEFRS 418
DB 344 LTQQAETFGIRHTLLYGLIELGYQKSDRVAA-----AGSFTYNNLPNP 386
QY 419 IPQV--DLIANARKGVGSHTVATENLDEFGYIGKSTFHPADGLSLIGGRLGHYKI-- 474
DB 387 VPVVLPTVPANATPSNYGLTNN-----ETVAMYAQDLIFKSPQWTYAGLR--YEVLK 437
QY 475 ESSEGGKT-----LHKASKTFGTGAGAVYDNLNDNNSLYLSLSQLYTP-----QTLDA 522
DB 438 QSRDLPKQODLSRTDK-PVSPRLGIVYHPVEALSLYASYSRQPLADSFYYYNSSA 496
QY 523 DGKLLKPRQGNQFEGYKGSYMDRLNARVS--FYRMKDKNAAAPLPPNNKTRRYAA-LG 579

```

497 -----LAPQSTTNYEIGAK--YDVSASASVSAALPDMKQTN-----LTSVDPATQALVPIG 545
580 KRMVEGVEITEISAVTPKQWQHAGYSLHSQIKTASNSRDDGIFL-----LMPKHSANL 633
546 TQTRGLELSFAGEVVKTVSWIASYAYLNG-VLTNFDKSNVSVSGNQPSLTPRHSGSL 604
634 WTYQVTPETLITGGVNA-----MSGITSSAGMHAGGYATPDMAAARFTPKLKLQIN 686
605 WVKHDLFNGPYVAGGLRAEGARFASQYVNT-----LPGYMTVDLGAGYR-SKHVDVTLN 658
687 ADINFRHYARV-GGANTFNIFGSERTWANLRY 720
659 LQNLFRAYIVSAHGAENYNLPGAPRMALLGVRY 693
RESULT 21
Q8YTX5 PRELIMINARY; PRT; 851 AA.
AC Q8YTX5; 2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALR2588.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AF003590; BAB74287.1; --
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR Receptor; Complete proteome.
SQ SEQUENCE 851 AA; 93901 MW; 5F93DD9E558D291D CRC64;
Query Match 12.2%; Score 460.5; DB 16; Length 851;
Best Local Similarity 23.6%; Pred. No. 7.1e-19;
Matches 173; Conservative 120; Mismatches 316; Indels 123; Gaps 25;
QY 17 VLAALSSVFAAQTADLETVHIKQGSYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQS 76
164 VLSAAPSQTQATPTPEEQPAATDQPIELVVTGQDQSVRYVTAATATKTDTPLRDIPQS 223
77 VSIITNQVDRNVDTFDQLARKTPGLRVLSNDDGRS--SVYARGY-EYSEYVNDGLPAQ 133
224 IQVTPROVEDQATQALNEALRNVSQVQ-SENSAGRTDRFVIRGFDPDFNSVRDGF-KE 281
134 MQSINGTLPLNLFADRVEMRGPSGLFDSSCGEMGIVNLVRKPTK--AFQGHAAAGFG 190
282 NTSVYRETANI--ERIEVLKGPASVLFQGLEPGGVINVTQVQREPFLLMGLEAGSYG 338
191 THQYKAEADVSGNSLNSDGSVGRVMAQT-VGASPRPAKNNRHETFYAAADWDINDPT- 248
339 F---FRPTVDNPSPLNDSKTLRLRVNAAVEISESRFDKETSFPFIAPVLAWEIGDRTS 395
249 -VLGAGYLYQQRHAPLYN-GLPADANNKLPSPQHFVFGADWNKFKMNSHVDVADLKHYF 306
396 LVFDLEYLKDOR--PFRDLGLVA-LGDRVAEIPFNRILGEPDSEVEDLRIGTRFEHGF 451
307 GNGGYGKGMVRDSDRADSNAFAGSKLGMKTPAGPGCNTADDKACAV-----GLGTEI 361
452 NDWKLRSAPRIIVSTQTSFVTEPGSL-----DEATGLLSRDFGVDRT 495
362 KQKALAFDASYSRPFPLGNTANEFVIGADYNRFRSTNEQGRTT-----LYARG 409

496 PDETVAFTDGLGFKTGAIEHELIVGDFNRTQNFVDDRSAPAPADIFNPVVLTRP 555
410 GLANERFSIQVDLIANARKGVRSHTVATENLDEFGIYKSTFFHPADGLSLIGGRL 469
556 NIESPPFRL-----FAADNI--GIYIQDQIKLAENLKLIGRY 593
470 GHYK---IESGEGKTLHKASKTKFTGYAGAVDLDNNLSLYLSQLYTPQTNLDADGKL 526
594 DFTNQSFSLIVDGVKFFDIDSOAFSPRIGIVYQPTPLSLYASYSRSFAQNFGIORDNSR 653
527 LKPRQNGFVGYKGSYMDRLNARVSFVRMKDKNAAP-----LNPNNKTKRYAALGKRV 582
654 IEPERGTOYEVGVRGEFLDGKLIASLAGYQITKTVATPDPADLNF-----IPUGEVR 707
583 MEGVETEISGAVTPKQWQHAGYSLHSQIKTASNSRDDGIFL-LMPKHSANLWTTTYQVTP 641
708 SRGIEFDIAGELAKGWNIIASAYTDAKI-TEDTNTDNEGNRLNRPVENSASITWTTVEL-- 764
642 ELTIGGVNAMSGITSSAGMHAG-----GYATPDMAAARFTPKLKL 683
765 -----QSCALQGLGNGVGLFFVGERQDLSNSFTVPGYTRTDALFYR-RDNWNI 813
684 QINADNIFNRHY 695
814 GLNFKNIFDVNY 825
RESULT 22
Q8X7W7 PRELIMINARY; PRT; 760 AA.
AC Q8X7W7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative outer membrane receptor for iron transport.
GN Z1026 OR ECS0883.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
DR EMBL; AE005261; AG55177.1; --
DR EMBL; AP002553; BAB34306.1; --
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 760 AA; 81956 MW; 60A294A347F7D502 CRC64;
Query Match 12.2%; Score 460; DB 16; Length 760;

Best Local Similarity 22.6%; Pred. No. 6.4e-19;
Matches 168; Conservative 109; Mismatches 333; Indels 134; Gaps 23;

```
QY 65 KIPASLREIPOSVSIITNQVDRVDFPDQARLTPGL-RVLSNDGRS-----SVYARG 119
Db 63 KFSRVAVATRTMTYISQVVKDQATNLTALAKVPRVGVAFPAENGNSTGDAIYRGG 122
QY 120 YEYS-ENYIDGLPAQMGINGTLFPLFADRYEVRGSPGLFDSGEMGIVNLVRKPT 178
Db 123 ADTSNSIYIDGI---RDIGSVSRDTFNTGEVVEIKGSPGTDYGRSAPLGSINMISKPR 178
QY 179 KAFQCHAAAGCTHKQYAAEDVSSLSNDSGVRRVWAQYVAGAPPAEKNNHETTYA 238
Db 179 NDGSDASASIGSAMFRGGLTDVNVIDDTTAVRLNMGKTHDGRKVNKR----- 232
QY 239 AADNDINDYLGAG-----YLYOORHLAPYNGLP-----ADANKLPS 277
Db 233 ---YGVASIAFGLGTANRLYLNTHTQHTPPDGGIIFTIGLPGYSAPSAGTATLHSGK 289
QY 278 LPOHVFVAGADWNNKFMNSHDVFAADLKHYFGNGYGVKGMRYSDRDADSNY--AFAGSKL 334
Db 290 VDTNHFYGTSDSDYDSTTDLTATMRFEHDINDVTTIRNTTRMSRVKQD--YLTMTALMGAS 347
QY 335 GAKTPAGRPG---CNTDADACAVGLGTEIKOKALADASTSRPFLGNTANEVIGAD 390
Db 348 NITQPTSDVNSWTSWTSRTANTK-----DVSNKILITQNTLTSTFYASIGHDVSTGE 399
QY 391 YNRFSTNEQGRITLYARGL---ALNEFRSIPQVLDLANKRGVRYSHVATENTLEDF 447
Db 400 FTREIQTN-----YGVNPTVTLPAVNITVHPDSSIHGGLTITNGANANGQT-----DTF 446
QY 448 GIYKSTFHPADGLSLIGGRIGHYKIE-----SGEG-----KTLHK 484
Db 447 AIVAFDTLQITRDFELNGIRLNDVHTEYDSATACGSGRGALITPAGVAKSPVTTTDT 506
QY 485 ASKTKFTCY-AGAVYDLNDNNSLYLSQLTTPQ-----TNLDADGLIKPRQG 532
Db 507 AKSGNLVMMKAGALYHLTENGVNYINVAVSQPPGANNFALAQSQSGNSANRTPKPOKA 566
QY 533 NOFEVGYGYSVMDRLNARVSFPRMKDAAALPNNKKTRYALIGRMEVGTELSG 592
Db 567 NTSEIGTWQVLDKRLTLTALLFRDITENEV---EQNDGDTYSQYKKRVEGEIISVAG 622
QY 593 AVTPKMQIHAGYSYLSQIKTASNSRDGIFLL--MPKHSANLMTTYOVTPELITGGVN 650
Db 623 NITPAMQVITGGTQOKATITKNGKDVADGSSSLPYTPHAFITMSQYATDISGAGAR 682
QY 651 AMSGITSAGMHAG-----GYATPDMAAAYFTPKLQINADNIENHYYAR 698
Db 683 YI-----GSMHKGSDGAVGTPAFTEGYVWADAKLGYRVNRLLDPQLNVVNLFDTDVAA 736
QY 699 VGANTEFNI PGSERTW--TANLRY 720
Db 737 INKSGRYHPGEPRTFLLTANMHP 760
```

RESULT 23

09HX3

```
ID 09HX3      PRELIMINARY;      PRT;      708 AA.
AC 09HX3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable outer membrane protein.
GN PA4837.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
```

```
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Loty S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004896; AAC08222.1; -.
DR HSSP; P06971; 1BY5.
DR InterPro; IPR002106; AAcRNA_ligaseII.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 708 AA; 79063 MW; 08FBE0D639CBB CRC64;
```

Query Match

12.2%; Score 459; DB 16; Length 708;

Best Local Similarity 22.7%; Pred. No. 6.7e-19;
Matches 172; Conservative 135; Mismatches 329; Indels 122; Gaps 25;

```
QY 22 SSSVFAQCT---ADLE--THIKQGRSYNAIVTEKNDYSEFATVGTGKIPASREIPQS 76
Db 16 SSPALMETESPALVELVLYAEAEKRAEPV-----QGYRANRTASARTDRIEDIPQA 70
QY 77 VSITNQOVQRNVDTFDPQLARKTPGLRVLSNDGRS--SVYARGYSEYENIDGLPAQM 134
Db 71 ISVPRQVLDLDSARLERALDFAGVSRQNNFGGLTFEYVNRGFTTSEYRDGFSANR 130
QY 135 QSINGTLENLPAFPRVEMRGSPGLFDSGEMGIVNLVRKPTPAQGHAAAGFTGHQ 194
Db 131 GYMNA--PDSATIRREVEIKGPASSLYGRGDPGVNLVTKKPAQERPARLHAASGMDR 188
QY 195 YKAEDVSGSLNSGVSRRGMAQTGASPRPAKNNHETFYA-----AADNDIN 245
Db 189 YRSTLDLNTPLDEBGDILLYR-----NLAVEDSGFRDYADGQRLVAPSIWQLD 239
QY 246 PDT-----VLGAGVLYOORHLAPYNGLPADANNKLPSPLOHVFV-ADWNNKFMNSHD 297
Db 240 PDTSLLYEAEVYRNKQVDFRGTVAPH-----NHLGSLPRSRFPGEPPDGKIDNNET 291
QY 298 VPADLKHYFGNGYGVKGMRYSDRDADSNVAFAGSKLGMKTPAGRPQGNATADACAVGL 357
Db 292 LQATLRHFNQWLSRLASHYKHGLD-GYASENSSLA-----ADGYSL 334
QY 358 GTEIKOKALADASYR---PRLGNTANEVITGADYNFRSTNEQGRITLYARGGLA 412
Db 335 RREYRYRDFEWHDSITQLDLGLDHTGSIHQILMGLEYERYH--NDE-----LI 382
QY 413 LNEFRSIPQVLDLA-NARKGVRY-----SHVATENTLEDFGIYKSTFHPADGL 461
Db 383 L--RSIPSRNPVALDIRRPYGOQPKPPFGRDDNH---EEVAMAMNLDDQIEFSKW 435
QY 462 SLIGGRLGHY-----KIESGE-GKTLHKAASKTFGYAGAVYDLNDNNSLYLSQ 513
Db 436 RGLGVRFRDRQDMNATRLNNGRFRETSSQGTQRAATPRIGVLYQATPEVGLPANSKS 495
QY 514 YTPQNLDAAGKLIKPRQGNQFVGYKSYMDRLNARVSFYRMKDN--AAAPLNNK 571
Db 496 FKPNGTDMAGKAPDPBEGRGYEAGVXLDLDLGGITGTLAFLHKKKRVLTADPSNPYQ 555
QY 572 KTRFAALGKRMEGVENEISGAVTPKMQIHAGYSYLSQIKTASNSRDDGIFLMPGSA 631
Db 556 QT---AGEARSGPDIQFSQQLTEQLIGAVAYIDAEVYKDNINARGSLNLVPPKSG 611
QY 632 NLMTTYQVTEPLLITGGVNA-----MSGITSAGMHAGGYATFDMAAAYFTPKLQ 684
Db 612 SLMGVEFRBEWMLGADAGAAVNVVGRBAGDSGSGFELPATTYVDLARPPLASNNATLG 671
QY 685 INADNIFNRHYAARVGAGANTFNPGSERTWTANLRYSP 722
Db 672 VVNVNLPDRRYER-SYNNVWVAPGEPRLNLTMSLTLYN 708
```

```

RESULT 24
Q9X7K9 PRELIMINARY; PRT; 726 AA.
AC Q9X7K9;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Outer membrane siderophore receptor precursor.
GN FHUA.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=384;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8401 PRL1;
RX MEDLINE=20244627; PubMed=10784041;
RA Yeoman K.H., Wisniewski-Dye F., Timony C., Stevens J.B., deLuca N.G.,
RA Downie J.A., Johnson A.W.B.;
RT "Analysis of the Rhizobium leguminosarum siderophore-uptake gene
RT fhua, differential expression in free-living bacteria and nitrogen-
RT fixing bacteroids and distribution of an fhua pseudogene in different
RT strains.";
RL Microbiology 146:829-837 (2000).
DR EMBL; AJ238208; CAB41037.1; -.
DR HSSP; P06971; 1QJQ.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR Receptor; Signal.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 726 AA; 79499 MW; BDEBF2A5CSAA0408 CRC64;

Query Match 12.1%; Score 458; DB 2; Length 726;
Best Local Similarity 23.8%; Pred. No. 7.9e-19;
Matches 189; Conservative 115; Mismatches 313; Indels 178; Gaps 31;

QY 12 MTAATVLAALSSVSFAAQTAD-----LETVHIKQORSYNAIVTEKNGDYSSFAV---TV 62
DB 22 VTTAIVLIGIAASPAASQASATASATALEPIVIQ-----GAASDSKADRTSVAANKSSA 76
QY 63 GTKIPASLREIPQSIVLIITNQVKDRNVDTFDQLARKTPGLRVLSN--DDGRSSVVARGY 120
DB 77 ATKINTPLVETPRSVVTEKEIEQRGAQSIIIEAVRYSAGVTTGPNFGDFPRFQIIFIRGF 136
QY 121 -----EYSEYNIDGLPAQMOSINGTLPLNLFAPDRVEMVRGSPGLFDS 163
DB 137 NITTVGYDRSLRQPIYNGMFRD--PYQLQ-----RVEVIKGPVSVLYGS 181
QY 164 GEMGGIVNLVRKPTKAFQGHAAAGFTGHQYKAEADVSGSLNSDGSVGRVMAQTVGAS 223
DB 182 GSPGLVNLKISKLPTEPIHEVGISYSTKRAQAMFDFGGPI-SEGN--DDFLYRIYVGLA 238
QY 224 PRPAEKN--NRHETFYAADWDINPD-----TVLGAGLYVQORHLAPYNGLPADANNKLP 276
DB 239 -RHGDNNFDIADRYFLAPSTFWKPDDEGTSFTLYG---LAQSDTDANVGAIITVDGKIL 294
QY 277 SLPOHVFVGADNWKFNKNSHDVPADLKHVFCNGGYGKVMRYSDRDADSNYAFAGSKLGM 336
DB 295 DIROS-----DPDYDYQVKVQQQIGYQFEHFDNGLTFTQNLYSHLDLRARYLVSSWTG- 350
QY 337 KTPAGRCPCNTADDKACAVGLGTETIKOKALAFDASYSRPFRGLNTANEFVIGADYNFRS 396
DB 351 -TVAHNRASSIRDE-----NNVFQVDNQLEAKFDGTGPLAHTMLFGLDY----- 392
QY 397 TNEOGRITLYARGGLALNEFRSIPQVDLIANARKVGRG---YSHTVATENLDEFGLYG- 451
DB 393 -----TNLQSNWGYGIGAVNSAFDFD-IANPTYGVSGATPAYDFIVADADMROGVIAL 445
QY 452 -----KSTFHPADGLSLIGGGRGLGHYKIESGEGKTLHKASKTK 489
DB 446 DOI EAGNWRNLGGRQTVWNTQTRDTTYPSPGLN-----DSEDVNKNA 487

```

```

Db 316 DIRS---DPDYDYQKVKQOQIGYFHEHFNGLFRONLAFYSHDLARLYGSWTG- 371
QY 337 KTPGRRCCTATTADKACAVGLGTEIKQAKALPDASYSPEFLGNTANEVIGADYRFRS 396
Db 372 -TVAHNRNASSIRDE-----MNVPQVDNQLEAKFPTGLAHMLFGLDY----- 413
QY 397 TNEGRTTLVARGGLANEFPSIPQVDLIANARKYGR-----YSHVTATENLDEGITYG- 451
Db 414 -----TNLOSNMGYGVGAVNSAFDP--IANPTYGVSGATPAYPFTVADAMROGVYAL 466
QY 452 -----KSTFHPADGLSLIGGRLGHYKESGEGKTLHKASKTK 489
Db 467 DOIENAMRFRNLGGRTQFVNVNTRDTTYSFGLN-----DSEVANKNA 508
QY 490 FTGAGAVYDLNNDNSLYLSQLYTPQTNLDADGKLKPRQGNQFEVYKGSY--DD 546
Db 509 FSWQAGALYLFEDNGIAFVSATSFDPVTNRASAGKILEPTEGQYELGVX--YQPGTD 566
QY 547 RLNARVSFYRHKDKNAAPLNNKRTYALGKRMEGVETEISGAVTPKMQIHAGISY 606
Db 567 ILTSAVA-YHIVERNKPVLPVNP--LTLAYDGLGEVTKGIELEKRAALADGLDIIMAYTY 623
QY 607 LHSQIKTASNSRDDGIFLLMKSHSANLWTTY---QVTP--ELTIGGVNAMSGITSSAGM 661
Db 624 NHESVTQGDNEGNTPAF--TPAHVASLMANYTFOETNPFNGLSGAGVRYVSEMWTDITAN 681
QY 662 HAGGYATF--DAMAAFRFTPLKQINADNIFNRHYARVAGANTFNIP----- 708
Db 682 TSKNPSFFYDASAAYDF-----GAIDKNVEGLTAAFNIRINADQRDVYCNEGF 730
QY 709 ---GSERTYANLRY 720
Db 731 CYLGGGRMTATLTKY 745

RESULT 26
O8U189
ID O8U189 PRELIMINARY; PRT; 747 AA.
AC O8U189;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE Ferichrome iron receptor.
GN ATU0409 OR AGR_C_718.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monke D.E., Kiteajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. St.,
RA Chapman P., Clendenning U., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-Y., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neeter E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Gurollo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Tarchouk O., Bpp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Guron J., Lomo C., Sear C., Strud G.,
RA Cielo C., Slater S.;

```

```

RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001).
DR EMBL; AE009012; AAL41430.1; ALT_INT.
DR EMBL; AE007978; AAK6224.1; -.
KW Receptor; Complete proteome.
SQ SEQUENCE 747 AA; 81287 MW; EBEDC551426F61C9 CRC64;

Query Match 11.9%; Score 450.5; DB 16; Length 747;
Best Local Similarity 25.0%; Pred. No. 2.3e-18;
Matches 195; Conservative 104; Mismatches 358; Indels 123; Gaps 33;

QY 2 GQFVSFRINTATVATYALSSVPAQADLETVHKGQSNVATYKNGDYSSPAVT 61
Db 24 GMFLSRALGTAAVVLGPPWG--HAQETTVLKQITVEGQGAENA--TGPVRYVANKXA 79
QY 62 VGTIPASIREIPQSVSLITNQVXKDN--VDTFQGLARKPFGLEAV--LSNDGSSVYAR 118
Db 80 TGSKTETETKALPOSVSIVGQEMDDRGAVTKIDEVLRTPGVTAPEPPTDPTDWFYIR 139
QY 119 GYEYSEYNI--DGLPAQMSINGTLPNLFAFDRVNRGSGFLPDSGEMGIYVNLVKR 176
Db 140 GFQATQTCVFLDGLNLTSYGRGFGQMAPYGLERVEVLKGPASVYIGGANBGIYQWMSKR 199
QY 177 ---PTKAFQ-----GHAAAGF-----GTHQYKAEADVSSGLN--SDGS--YRGR 214
Db 200 AQDTPVRETEIGINNFNAPFGFDLGDKVDABGVWK--YRVTGKVSQGDNYTDESDLRGF 258
QY 215 VMAQTVASPRPAKNNRHEFFVAAAWMDINPVLGNGYLYQQRHLAPYNGLPADANAK 274
Db 259 IMPQ-ITFEP--DAQTSATLYGYFSALD-----QVHANGFL-----PYVGTVVDA--P 302
QY 275 LPSLPQHVFVG--ADMNFKKNSHDVFPADLKHFFNGGNGYGVKMGKYS-- 320
Db 303 FGKLDRAKAFYCEPDIDNGRYQSVNGYDVSHFPNGKISQMARVGHLYKHETGPPYRGM 362
QY 321 RDADSNYAFAGSKLGMKTPAGRPCCN--TADDKACAVGLGTEIKQAKALPDASYSRPFRL 378
Db 363 ANADAN-----GQPIIDPITNDYMLTRFGYDGLSKVDSFGVGNRIEQQFDT 408
QY 379 GNTANEVIGADYRFRSTBQGTITYANGGLANEFPSIPQVDLIANARKYGR----- 434
Db 409 GAVNHSPLFGIDYKYRYLDVOVQACGSAIICAL-----KPYVGSITOG 450
QY 435 -----YSHVTATENLDEGITYGKSTFHPADGLSLIGGRLGHYKIESGE-----GKTLHKASK 487
Db 451 TNEFYADNITYQ--QQGIYAQDQLRFGDGLVTLNCRYPYVDTLNNRLPAGVSRSSND 508
QY 488 TKFTGAVAGVYDLNNDNSLYLSQLYTPQTNLDADGKLKPRQGNQFEVYKGSYMD 545
Db 509 DALSGRAGLAVFPDNGLTTPYVSATFPNPLDITLADGTPASPEBGHQFEGAIKTEPSPFD 568
QY 546 DRLNARVSFYRHKDKNAAPLNNKRTYALGKRMEGVETEISGAVTPKMQIHAGISY 605
Db 569 GSITASV-FKLVDK-NAIVSYTAGGVVTS--GQFQVSVSTFTELEKANKLNDENWALASYS 625
QY 606 YLHSQIKTASNSRDDGIF--LLMPKSHSANLWTTYQVPE-----LTIGGV--NAMSGITSS 658
Db 626 YTTDIDITKDNANPLIGSPVIVPAHTASLWVDYAFDTETEGSLIGGAVYQKGSWADAA 685
QY 659 AGHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVAGANTFNIGSERTWTANL 718
Db 686 NTLKRVSDAAVFDAIARE-KNDWMTASVNVANVPDKEXVYKSCAGSVCGW-GDSRTITLKL 743

RESULT 27
O8UK13
ID O8UK13 PRELIMINARY; PRT; 819 AA.
AC O8UK13;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE TonB-dependent receptor.

```


QY 249 VLGGYLYQOHLALYN-GLPADANNKLPSPQHVFGADWNKFRONSHVYFADLKHYFG 307
 Db 413 TLOEFLYNDER---PFDGFLAFEGGIDT-PLERFFGEPDVRKVEIGISYREHNFN 468
 QY 308 NGGKGVKVMRSDDADSNVAFAGSKLGMKTPAGPCGCTADDAKAVGLTEIKOKLA 367
 Db 469 DNWKIRNAFRYQSSDT-FDYRAEPRLNETTGILTRNRSNDDYEENYTLQTDVVGK--- 524
 QY 368 FDASYSRPFRLGNTANEFIVGADVNRFRSTNEQGRITLYANGSL---ALNEFRST- 419
 Db 525 -----FITGSIINHTLLPFDVDLARFISGQTOSHP-----GVTTPSINFPYNAIP 571
 QY 420 -POVDLIANARKVGVGSHVATENLDE---FGYKSTFPADGLSLIGGRLGHYKIE 475
 Db 572 RFGDELTLN-----VYRNNQDSTGIGIFLQNGIAFADNLKLVGSRFPYVQGN 620
 QY 476 SGEKTLHKASK--TKFTGYGANYDLNDNNSLYLSQLYTPQTNLDADGKLKPRQGN 533
 Db 621 STDLRDSESSEGRYDAFPRIGIYQPIEPISTLYASYSQSPQNFQTRVDSIIEAERG 680
 QY 534 QPEVYKGSYMDRLNARVSFYRMKDKNAAAP--LNPNNKTRVAAAGKRVMEGTEIS 591
 Db 681 QYEVGKGEFIDGLAATLAAYHTTKSNIAATDLDNDP---FLPIGEQNGIELNVA 736
 QY 592 GAVTPKMQIHAGSYLSHQSITKASNSRDGIFLMPKHSANLMTTYQV---TPELTIGG 647
 Db 737 GEISPGMNVIASYSHIDAEIRNDGLQGNRPANVPFNTASFTWYIELQOQDLOGLGRL 796
 QY 648 G---VNASGITSAGMHAGYATPDMAAYRFTPKLQIINADINFRHYARVAGANT 704
 Db 797 GLFYVGRDGDSSNTYI-IPGYLRTDA-ALYYKRDNRAGINIQNLFEKHYL---GANF 851
 QY 705 FNI---PGSERTWTANLRYSF 722
 Db 852 GRVAIEFGALPVTYIGSFVTF 872

RESULT 29
 068590
 ID 068590 PRELIMINARY; PRT; 753 AA.
 AC 068590;
 DT 01-AUG-1998 (TReMBLrel. 07, Created)
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
 DE HYDROXAMATE-type ferrisiderophore receptor (iron transport protein
 DE Ftu) (Probable outer membrane receptor for iron transport).
 GN P1UA OR F1U OR PA4514.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxId=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=96210657; PubMed=8633080;
 RA Ochman U.A., Vasili M.L.;
 RT "Gene repression by the ferric uptake regulator in Pseudomonas
 RT aeruginosa: cycle selection of iron-regulated genes.";
 Proc. Natl. Acad. Sci. U.S.A. 93:4409-4414 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RA Ochman U.A., Vasili A.I., Johnson Z., Vasili M.L.;
 RN Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Schwan W.R., Barker L., Brody L.L.;
 RT "Differences in sensitivity to PA-1806 among iron transport mutants of
 RT Pseudomonas aeruginosa compared to Escherichia coli.";
 Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Gollity L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 Nature 406:959-964 (2000).
 RL Nature 406:959-964 (2000).
 DR EMBL; AF051690; AAC06215.1; -
 DR EMBL; AF276976; AAF87743.1; -
 DR EMBL; AE004665; AAG07902.1; -
 DR InterPro: IPR000531; TrnB_BoxC.
 DR Pfam; PF00593; TrnB_BoxC; 1.
 KW Receptor; Complete proteome.
 SQ SEQUENCE 753 AA; 82336 MW; 4DD5430DC9D6514B CRC64;

Query Match 11.7%; Score 441.5; DB 16; Length 753;
 Best Local Similarity 22.9%; Pred. No. 7.8e-18;
 Matches 186; Conservative 129; Mismatches 318; Indels 179; Gaps 37;

QY 15 ATVLAALSSVFAQTADLETFHIKQORSYN-----ALVEKQGDYSSFAV--TVG 63
 Db 17 ASAIGVATITAAQAQADEA---GQKTDKRVLSLDAATYGEQODETTTVNDRSAS 73
 QY 64 TKIPASIREIPQSVSIITNQOVKDRNVDTFDQARKTPGLRVLSNDRGRS---VYARGY 120
 Db 74 KKYAPALDTPKVTYVLPQVIKDTGALTLADALRTTPTGTFGEGGNPAGRPFRGF 133
 QY 121 E-YSEVNIIDGLPAMQOSINGTLPLVAFPDVEVWRGSGLPDSSGEGGIVNLVKKRPTK 179
 Db 134 NAESDTFLDG---MRVVASQTRVEFVNEQIEVSKGSGASYTGSGGSGSINTLISK--TA 187
 QY 180 AFOGAAAGF--GTHKOYKAADVSGSLNSDGSVGRVMAQTVGASPRAPKKNRHETFY 237
 Db 188 KQDNFTDAGFTWGSDDQTRRTLLDVRNMGIDMAAFRLNM-----KDAHV 232
 QY 238 AAAD-----WDINPDTVLG-----AGLYQORHLAFNGLPADANKL--PSLPQ 280
 Db 233 AGRDEVSVSRMGVAPVTFPFDPTFRATLSYHLSTDMDPDYGLPLTNVRSKANKPKPA 292
 QY 281 HV---FVG---ADMKFKKNS-----H-----VFADLKHYFGN--GGYG 312
 Db 293 SVPRDNFYGLKDRDYKSTDSGTFRIEHLNDNLTLNSSTRLVRTLLDIVSNPDDSRG 352
 QY 313 KVGKRSYDRDADSNVAFAGSKLGMKTPAGRPGCNTADDAK-----CAVGLTEIK 362
 Db 353 NVANGIYIRAKS-----KUSTSKGMVNGQTDLKANPFTGFIKHTLVTL 396
 QY 363 QKALAPDASYSRPFRL-----GNTAN-EFVIGADVNRFRSTNEQGRITLYANGGLALNE 415
 Db 397 --EFSYEDVNRPRYAITSGGAGNTCNARLASGDCSLNRPFGDNWTSITDGLAYTD 454
 QY 416 FRSLPOVDLIANARK-----GVRGSHVATENLDEFGYIGYSTFPADGLSLIG 466
 Db 455 TDTKTSAAVYFTDLKLSQEWELNLGLR-----YDFD--KSSGYQTAGR-- 498
 QY 467 GRLGHYKIESEGEGKTLHKASKTKFTGY-AGAVYDLNDNNSLYLSQLYTP-----Q 517
 Db 499 GPAGYFKRE-----NSHFMYQYGLVYKPAKNGSIYLAWSSTSNPTEGEGQ 548
 QY 518 TNLADGKLKPRQNGQFEVYKGSYMDRLNARVSFYRMKDKNAAAPLNPNNKTRVYAA 577
 Db 549 ADISVGNNGLDPERNRNLLEIGTKWAFFDALSLNALFR-TDKTNARVASP-DVSTLYVL 606
 QY 578 LGRVWEGYETELSGAVTRKQVHAQSYLSHQSITKASNSRDGIFLMPKHSAN--LM 634
 Db 607 DGEQVQGVELGNGKLTERRKVFQGYTYLDSIRKSYKSGG--KNMPQTAQNNFTLM 664
 QY 635 TTYQVTPPELLIGG--VNAMSGITSSAGMHAGYATPDMAAYRFTPKLQIINADNIF 691

[illegible]

QY	389	ADY--NRPRSTNEQG-----RT-----TLYARGGLALNFRSIPQVDLIAN	427
Db	404	KFPANGFDLHNPFGTWNGPTDKDSARNRTELEVKALYAFDTIALDE-----RWDLSLG	458
QY	428	AR-KGVRGYGHTVATENLDEFGYIGKSTHFPADGLSLGGRGLGHYKYESGGKTLHKAS	486
Db	459	LRWDWIDGTSRSTPS-----GKPT-----VRDSSDG-----	485
QY	487	KTKPTGVAGAYVDLNDNNSLYLSQLYTPQTN-----LDADGKLLKPRQGNQFV	537
Db	486	--KLSTRAGLVFKPLENGRVYFSGTFSNPSAHLVTTSGVTEATGG-LAPEKNSYVEL	542
QY	538	GKGSYMDRLNARVSFVRMKDKNAAPLNPNKKTYYAALGKRVMEGVETEISGAVTPK	597
Db	543	GTKWEFLGRRLLEDAALFOVKKEDARERLADGS---YVLAGEQRVGRLESASGKLNEH	598
QY	598	WQIHAGYSLHSQIKTASNRDDGIFL-LMPKISANLWTTVQVTPELTICGGVNAMS--G	654
Db	599	WDLFGTYIYLDSETLKSNPQREGQALGNTPPRSLSLWSTVELPEGVTLGYARYYSORN	658
QY	655	ITSSAGMIAGGYATFDAMAARFTPKLQJLNADNINFRHYIARV	699
Db	659	VTSVONGKLDAYVHNHMLGYQATRDCLKLQNLNLFDAKVERV	703
RESULT 31			
QY1422	QY1422	PRELIMINARY;	PRT; 732 AA.
AC	Q91422;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-OCT-2001 (TrEMBLrel. 18, Last annotation update)		
DE	DE	Probable TonB-dependent receptor.	
GN	PA1322.		
OS	Pseudomonas aeruginosa.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;		
OC	Pseudomonas		
OX	NCBI_TaxID=287;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 15692 / PA01;		
EX	MEDLINE=20437337; PubMed=10984043;		
RA	Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,		
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,		
RA	Gardner R.B., Goltry L., Tolentino K.R., Kas A., Larbig K., Lim R.M.,		
RA	Brody L.L., Coulter S.N., Folger K.E., Wong G.K.-S., Wu Z., Paulsen I.T.,		
RA	Smith K.A., Spencer D.H., Hancock R.E.W., Lory S., Olson M.V.;		
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;		
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an		
RT	opportunistic pathogen."		
RL	Nature 406:959-964(2000).		
DR	EMBL; AE004561; AAG04711.1; --		
DR	InterPro; IPR000531; TonB_boxC.		
DR	Pfam; PF00593; TonB_boxC; 1.		
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.		
DR	Receptor; Complete proteome.		
KW	RECEPTOR; 732 AA; 80835 MW; C2171C0FDC409FBB CRC64;		
SQ			
Query Match 11.5%; Score 432.5; DB 16; Length 732;			
Best Local Similarity 23.3%; Pred.No. 2.5e-17;			
Matches 178; Conservative 123; Mismatches 299; Indels 165; Gaps			
QY	17	VLAAGSSVFAAQT-----DLETVHKGQRSYNAIVTEKNGVYSFVAVTGKIPASL	70
Db	22	LMAACSSPALAEADAARKDDPLELGADTVTGEQASSRVERSASAKYA-----VP--L	70
QY	71	REI PQSVSIITNQVQKDRNVDTFDQLARTPLGLVLSNDDG---RSSVYARGY-YSEVN	126
Db	71	LDTQTQTVTVVQKVIQEQNALSLRQVLSNVSGITTFNAGEGGGGSGDSINIRGFSANANWQ	130
QY	127	IDGLPAQMQSINGTLFNLFAFDVEVMRGPSGLFDSGEMGGIVNLVRKPTKAFQGHAA	186


```

Db 131 VDLGRSAQT---SRSDLFNLEAVEVITKGRNSVFGAGTSGSINMWSKQPKAEFTHLG 187
Qy 187 AGFTGHKOYKAEDVSGSLNSDG---SVRGRVMAQTGASPRPAEKNNRHETFYAAADWD 243
Db 188 AGLGTRVYRLTLDTNQLLEGSTAFRLNMAHNDVAGSEQIDKQR-----WG 238
Qy 244 INPDVVLG-----AGYLVOQRHLAPYNGLPADANNKLSLPQHVFVGADWNK---K 292
Db 239 IAPSLTFGLDPTRTLTLTSFYHQRDNDLPDYGVPALNGRKLGVSHRDYFG--WRMLDEER 296
Qy 293 MNSHDFADLKHVFGNGYGVKVMRYSDDRAD-----SNVAFASGKLGMKPPAGPGCNT 347
Db 297 IDNVAATLKLEHDESDDQLOLRLYSHLRDVTYSASHNOKGLPPRGYRLP--PQAYG 355
Qy 348 ADDKA-----CAVGL-----GTEIKOKALAFDASYSRPRPLAGTANEPYIG 388
Db 356 RDSKTRMWINQTNLGRFDTGLAHTLLIGFELSRE-----TYR-----TYSYMLG 403
Qy 389 ADY--NRFRSTNEQG-----RT-----TYARGGLALNEPRSTPQVDLIAN 427
Db 404 KFYANGFDLHNPFGYNGPPTDKRDSARNRTELEVKALYAFDTIALDE----RWDLSLG 458
Qy 428 AR-KGVREYSHVATENLDEFIYVGSSTFHPADGLSLGGRLGHYKLESSEGKTLHKAS 486
Db 459 LKYMIDTSTKSTBS-----GKPT-----VRADSSDG----- 485
Qy 487 KTKFTGAYAVYDLNDNNLSLYLSQLYTPQTN-----LDADGKLKPRQNOPEV 537
Db 486 --KISTRAGLVFKPLENRVYFSYGTSTFNPASHELVTTGSGVTEATGG-LABEKSTYEL 542
Qy 538 GYKSYWMDRLNARVSFFRMKDKNAAAPLNPNNKTRRYAALGKRVMEVETEISGAVTPK 597
Db 543 GYKMEFLGRRLLEALALQVKKEDARERLADGS-----YVLAGEQVRGLSASGLKNEH 598
Qy 598 WQIHAGSYLHSQIKTASNSRDDGIFL-LMPKHSANLMTTYQVTELTIGGVNAMS--G 654
Db 599 WDLPAITYTLDESETLKSSNPREGALANTPRSLTSMSTYELPBGVTLGYGARYVSQRN 658
Qy 655 ITSSAGMHAGYATFDAMAAYRFTPKLKIQINADNIFNRHYARY 699
Db 659 VTSVONCKLDAYVWVNMMLGYQATPDLTLQTLNDLFLPKAYVERV 703

RESULT 32
Q9FOF9 PRELIMINARY; PRT; 755 AA.
AC Q9FOF9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Putative ferrichrome iron receptor PhuA.
GN PHUA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21064758; PubMed=11136137;
RA Galindo M.A., Day W.A., Raphael B.H., Joens L.A.;
RT "Cloning and Characterization of a Campylobacter jejuni Iron-Uptake
RT Operator.";
RL Curr. Microbiol. 42:139-143(2001).
DR HSSB; AF241226; AAC47637.1; -
DR HSSP; P06971; 1BY5
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KM Receptor.
SQ SEQUENCE 755 AA; 83851 MW; C4140C3D258202E0 CRC64;

```

Query Match 11.4%; Score 432; DB 2; Length 755;
 Best Local Similarity 23.9%; Pred. No. 2.8e-11;
 Matches 179; Conservative 116; Mismatches 331; Indels 124; Gaps 31;

```

Qy 15 ATVLAALSSVFAQTADLETVNHKGGRSYNAIVTEKNGDYSSFA--VTVGTKIPASIRE 72
Db 18 AVVLAATCHGVQAD---DQVTLPSQVTGELBRPQRPDGYAKESLTLTSTPLSE 74
Qy 73 IPQSVITITNOQVADRNVDTFDQARKTPGLRV--LSNDGRRS--VYARGYESEYNI- 127
Db 75 TPRSVSVTRKRIEQKQSLVDVLGYVPGIFAPPPAAGDGLAGDLFFIRGFNATDYGVG 134
Qy 128 ---DGLPAQOMSINGTLNLFAPDRVEVMRGPGLFDSSEMGYVNI.VRRKPTAPQGH 184
Db 135 LLRDLRVQGNRYD-TISEPYGLERVEFRGPTSLYGENAPGLGVNLSVRKPTATQGE 193
Qy 185 AAAGFTGHKOYKAEDVSGSLNSDGSVRGRVMAQTGASPRPAEKNNRHETFYAAADWDI 244
Db 194 VQLGYSNDRQRLNLDVSGPLNDSGNVIGRV--MVGSEAPQVDYHDVDDIYIAPSLTL 251
Qy 245 NPD-----TVLAGYLYVOQRHLAPYNGLPADA---NNKLPQLPHVFGV-ADWNKFKMNS 295
Db 252 NFDVNSLTLLST---YQKDRTKLELGLPAGTLLRNPNKIDKDTFLGNKDMWTFEREV 308
Qy 296 HDVFAADLKHVFGNGYGVKVMRYSDDRADSNVAFASGKLGMKTPAGRGCTADDAKAV 355
Db 309 WTLGYEFTFRNDWMQFRNSRYMQSRIQRNEIWPGA-----LNNA----- 349
Qy 356 GLGTEIKOKA-----LAFDASYSRPRPLAGTANEPVIGADYNRPRSTNEQRTTLY 406
Db 350 GFGTNTVQAYDRNKSTITYSGLDNQBEKFTDGLAQHTLVAGASFR--TSRND-----W 403
Qy 407 ARG-GLALNEFRSI---PQVDLIANARKGVGYSHVATENLDE--FGIYKSTFHPAD 459
Db 404 SAGFGPSINVENPWTSEPTTPF-----TIONSQLOQWAGLYSQLSKYEN 450
Qy 460 GLSLIGGRRLGHYKLESSEGKTLHKASKTK-----FTGYAAVYDLNDNN 504
Db 451 WIFLLGG-----RFDVDSQYRNKFDATSPVSNASDLDYTDSDFTWQAGVMYQFENG 504
Qy 505 SLVLSLSQLYTP-OTNLADADGKLKPRQNOPEV--YKSYWMDRLNARVSFFRMKDKN 561
Db 505 SPVYSYSFFPVQQTTSASGP-LDPTTAEYEGVGLKTEPRGNNTMPTAAV-FDLRKND 562
Qy 562 AAAPLNPNNKTRRYAALGKRVMEVETEISGAVTPKQIHAGSYLSQI-KTASNSRD 620
Db 563 VYFPAQIND---YRQVESRSKGELEINSDISANLVTAAYTYTDARIKDAPGSLLE 618
Qy 621 GIPFL-MPKHSANLMTTYQ-----VTBELTIGGVN-----AMSGITSSAGMHAGYATF 669
Db 619 GNMQTVGVRNQASIMANYRFLDGLALRGIGVGRHFFHSTFAYTAESLYGKLDIGDYTLV 678
Qy 670 DAMAAYRFTPKLKIQINADNIFNRHYARY 699
Db 679 DAALGYDIDENWSVDLNKKNFFDQGVGRGL 708

RESULT 33
O86424 PRELIMINARY; PRT; 732 AA.
AC O86424;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Ferrichrome iron receptor.
GN PHUA.
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K4;
RX MEDLINE=98348446; PubMed=9683481;
RA Killmann H., Herrmann C., Wolff H., Braun V.;
RT "Identification of a new site for ferrichrome transport by comparison

```


Qy 669 FDMAAARFETKLCQINADNIFNRHYYAVGANTENIPGSERTWTANLRYSF 722
Db 809 TDAEIFYK-RDNWRAALNFKNLFDTKYYE---SOSFFIVAPAFVLTGVSEFF 858

RESULT 35

Q926C6 PRELIMINARY; PRT; 724 AA.
AC Q926C6; TREMBREL. 19, Created)
DT 01-DEC-2001 (TREMBREL. 19, Last sequence update)
DT 01-MAR-2002 (TREMBREL. 20, Last annotation update)
DE Putative ferrichrome-iron receptor precursor protein.
GN FH02 OR R02269 OR SMC01657.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Rhizobiaceae; Sinorhizobium.
NCBI_TaxId=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batur J.,
RA Bolstead P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Golliau A., Kahn D., Kiss E., Lelaure V., Maury D.,
RA Pohl T., Portetle D., Puhler A., Purnelle B., Rampeger U.,
RA Renard C., Thebaud P., Vandebol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021".
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591790; CAC46848.1; -.
DR InterPro; IPR000531; TonB_boxC.1.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 724 AA; 79240 MW; CB919FBBE9794A9C CRC64;

Query Match 11.2%; Score 423; DB 16; Length 724;
Best Local Similarity 23.5%; Pred. No. 9e-17;
Matches 189; Conservative 112; Mismatches 361; Indels 122; Gaps 27;

Qy 5 MSVRINMTATVLAALSSVFAAQTDLFVHKGQSRNAYITKNG-----DYSSF 58
Db 9 LPLRLLAGTSALALVATA--QAQAEQETVSGNDSTALETLVNGSGGVTAEAGVGT 66
Qy 59 AVVTGTIPASIREIPQSVSIITNQVDRVVDTFDQARTPGARV--LSNDGRSSVY 116
Db 67 SSATGAKIDTPELETPOSISTVTEQQLDKRNPOTLLETLAYTPGTRVGAYGDFDPDAF 126
Qy 117 ARGYE--YSEYNIDGL--PAQMOISNGTLPLNFAFDRVVRGSPGLFDSGEMGIYVL 172
Db 127 VRGFDVTVYSGVFRNLQPAVDSIFKNEP--YGLEGSILRGSSALYKATGAGGLYVL 184
Qy 173 VRKPTAFQGHAAAGFTGHKOYFAEDVSGLSNDSGVRGVW-----AQTVGASPR 225
Db 185 ITRKPTEDTLREYOVGYGSHDRYQGFEDFSGPVNENDPVYRLTGLRLADDTQGVGLAD 244
Qy 226 PAKGNHETFFYAADMINDP-----TVGAGLYXOORHLAPFNGLRPADANNKLPSLPQ 280
Db 245 RA-----YIAPATFWKDEDETKLTVLGE--YSTTN---SGGATYYNDPLTGEAT 289
Qy 281 HVEVG-ADWVKFKKNSHDFADLKHVFGNGGYGKVMRYSDRADSNVAFASKLGKMTF 339
Db 290 DIFAGNDFPDSVQKQKRGVGEFEHRLNDTFVRQNAKRVSTLNTIDADMAFAYAP----- 343
Qy 340 AGRGCTADDKACAVGLGT-ETQKALAFDASVSRPRLGNTANEVIGADYNRFSTN 398
Db 344 -----NAAPPTLLDSSAGTYDERLTFVLDNQLEAKFDTGALHTHTLLAGDVYTKLFR 397
Qy 399 EOGRTIYAGGLALNFRSIPQYDLANAKRGVSHVATENLDEFGIYGSTHRA 458
Db 398 LDGKGVBPDLDTKNPQGRVDAIDP-----NRTYVDQWQLGYTLDQIR-Y 444

Qy 459 DGLSLIGGRGLGHYKISGECKTLHKASKT-----KFTGYAGAVYDLNNSLYLS 509
Db 445 DAMTLTVGR---YDWMSTDTDTMDLTDLSLTVYSQDKERSGRLGLTYQDFGLAPYIS 501
Qy 510 LSQLYTPQTNLDAD-GKLLKPRQGNQFEVYKSGSYMDRLNARVSFRMKDKNAAAPLNP 568
Db 502 YSTAFAPNAGINKETNOPFKPTEBEOEIGVKYLLPNSNTLITLALFNIDKNGLY-LEA 560
Qy 569 NNNKTRVALLGKRVMEGVEETISGAVTPKMOIHAGYSYLSQITASNRDGIFFLMPK 628
Db 561 SCDTALIVQKRLKASRGEIENANSLDNGISLISAYTVDKTIQGSRTGNTVSSAPQ 620
Qy 629 HSANLWTTYQVTPR-----LTIGGVNAMSGITSSAGMAGY-----ATFDMAAARF 677
Db 621 HMASIMAHYTL-PEDGPFYFSLGGARF---VGSSTGNQDNTKNSRVLFDASVGVDF 676
Qy 678 T-----PKLQINADNIFNRH-----YAVVGANTFNIGSERTWTANLRYSF 722
Db 677 AAIQKYEGLHLQVNAATNLPDRRAVCTAGYCHRDQ-----RTVIGSIRYMW 724

RESULT 36

Q82908 PRELIMINARY; PRT; 696 AA.
AC Q82908; TREMBREL. 20, Created)
DT 01-MAR-2002 (TREMBREL. 20, Last sequence update)
DT 01-JUN-2002 (TREMBREL. 21, Last annotation update)
DE Periloxamine B receptor.
GN STY0396.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxId=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L.T., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jørgels K.,
RA Kosch A., Larsen T.S., Leather S., Moutie S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL627266; CAP08819.1; -.
DR InterPro; IPR000531; TonB_boxC.1.
DR Pfam; PF00593; TonB_boxC.1.
KM Receptor; Complete proteome.
SQ SEQUENCE 696 AA; 77130 MW; E3FCEA55PD723FF CRC64;

Query Match 11.2%; Score 421.5; DB 16; Length 696;
Best Local Similarity 23.1%; Pred. No. 1e-16;
Matches 175; Conservative 121; Mismatches 353; Indels 109; Gaps 25;

Qy 12 MTATVLAAL-----SSSVFAAQTADETVHKGQSRNAYITKNDYSSFAVTVG 63
Db 1 MFATTRALLIGALIGATFPFLAQETTKNDV-----IVT-----SPVSGA 43
Qy 64 TKIPASIREIPQSVSIITNQVDRVVDTFDQARTPGRLPVLSNDRGSS-----VYARG 119
Db 44 TRLATPDIETPOSVSITTRQOFEOGATSVRAQVSYTPG--VYSNQIGASNRFDYIVLRG 101
Qy 120 YXESEVY---IDGLP--AQMOISNGTLPLNFAFDRVVRGSPGLFDSGEMGIYVLVR 174
Db 102 FSDGSLDNVTLIDGKMMGDTNNSHSLVVDWFLIEDLEVAGPASVLYGRSSPGGIYSLTS 161
Qy 175 KRPTKAFQGHAAAGFTGHKOYKAEDVSGLSNDSGVRGVWQTVGASPRPAKKNRHE 234
Db 174 KRPTKAFQGHAAAGFTGHKOYKAEDVSGLSNDSGVRGVWQTVGASPRPAKKNRHE 234

```
Db 162 RKPAPDAGGEVKLFAGNNQNRGAAPDVTGPDLDNERNVAAALSGWTRYADSQFTPLKEERY 221
QY 235 TFYAAADWDINPDTVLG-AGYLYQ-----QRHLAPYNGLPADANNKLSLPQH 281
Db 222 ALMPSLTWRTITDRDLDMAYLHRDPQGGSHGLPYQGTWVPYNG-----GKISNT 272
QY 282 VVFVGA--DWNKFKMNSHDVADLKHFGNGGKYGKVMRYSDRDADSNYAFAGSKLGMKTPA 340
Db 273 FPEGEDDYDKYDRRENKVNNGVNIHLEFDNGMSVRQKRLYHTKVTLLNQVYAAGWLN----- 327
QY 341 GRPGCMTADDKACAVGLGTEIKOKALAFDASYSRPRNGTANEFVIGADYNRFRSTNEQ 400
Db 328 -----ETALNRGYS-----GSGEKMSAIALDNQDGSVDTGAINHRLVGLDYQ-----DRS 374
QY 401 GRITLYARGGLALNEFRSIPQVDLIANARKVGRVGSHTVATE--NLDFEIGYKSTFHFA 458
Db 375 NHTTGY-----YGAFFPIDAFNPVYGAQPDYITLRSREKHLKQGTGYLQDQM-SW 424
QY 459 DGLSLTGGRLGHYKESGECKTLHKA-----SKTKGTGAGAVYDLNNDNSLYLSLSQLY 514
Db 425 DRWRFTLGGRYDRSVSNID--KLHDSRSLDKNNVSTRAALLYLFDNGVAPYLSYSTAF 482
QY 515 TPOTNLADAGKLLKPRQGNQFEGYKGSYNMDDLNRARVSPYRMDKNAAAPLNPNKKTR 574
Db 483 TSTSFADENGVLEPKMGKOWEAGVKEPLGNSQFSAAVYRINQINATKEEPTDP--- 539
QY 575 YAAALGRKVMGVEVETISGAVTPKQWIKHAGSYLHSHQIKTASNRDDGIIFILMPKHSANLW 634
Db 540 YRSIGETESKGELEAISHLSDSVRLQAAVYTDIIRYKSSPQEQGKRAVYAPRNOASAW 599
QY 635 TTQVTPTE-----LTIGGVNAMSGITSS-AGMHA-GGYATFDMAAAVRFTP-----KLKQ 684
Db 600 LSYDVKSGLLEWLTGSGIRYVNGVTSRDLNTHLTPSYTLVDMVVGVDLSIGLGLSAQ 659
QY 685 INADNIENRHYARVGCANTFNIPGSRBTWTANLRSYF 722
Db 660 LNVNLTDKRYVACNSL-SYCYFGEARSIVGVSNAF 696

RESULT 37
Q8YU26 PRELIMINARY; PRT; 853 AA.
AC Q8YU26;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALR2185.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840; Kuritz T., Sasamoto S.,
RA Kaneko T., Nakamura Y., Wolk C.P., Ishikawa A., Kawashima K., Kimura T.,
RA Watanabe A., Iriguchi M., Ishikawa M., Matsumoto M., Matsuno A., Muraki A.,
RA Kishida Y., Kohara M., Takazawa M., Sugimoto M., Yamada M.,
RA Nakazaki N., Shimpo S., Yasuda M., Tabata S.,
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003588; BAB73884.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete Proteome.
SQ SEQUENCE 853 AA; 94548 MW; 67EA77C09E049437 CRC64;

Query Match 11.1%; Score 420.5; DB 16; Length 853;
Best Local Similarity 23.1%; Pred. No. 1.6e-16;
Matches 187; Conservative 117; Mismatches 327; Indels 177; Gaps 30;

QY 3 QFMSVFRINMTAATVL--AALSSS-----VF--AAQTADL-----ETVHIKQRSYN 45
```

```
Db 135 QDANTIRSVIGETILTAEFLFDSDEGLIFGLTSVTSASLPQPQETQSEQSPAEEPIE 194
QY 46 AIYV-EKNGDYSSFAVTVGTKI PASLREIPOSVSIITNQVKDRNVDTFDQLARKTPGLR 104
Db 195 LVVTGEKDGQVQVNTATV-TRTNTPIIDIPOSIQVPPQVLEDDQOITRVDDALRVPGVL 253
QY 105 VLSNDGSRSSVVARGYEYSEYNI--DGLPAQMQSINGTLPNLFAPDRVEMVRGSGFLFDS 162
Db 254 GSTNAFIGNQITIRGFTSNLPILLRDPRIYE--NFSQFETSNIERIEVLKGPASVQVG 310
QY 163 SGMGGIVNLVRKPTKAFQGHAAAGFQTHQYKAEADVSGSLNSDGSVGRVMAQTGCA 222
Db 311 QLDPCGVNLVTKPLSEPFVEIOAQFGSGLIRPSFVSGPLTDDGKLLYRLNA-TYQR 369
QY 223 SPRPAEKNNRHETFAAAD--WDINPDTVL--GAGLYQQORHLAPYNGLPADANNKLPSL 278
Db 370 BEGPRDFNTETEFPIAPSLTWKISDRTNVDFSLYLOSTR---PFDTSLVAFGRSVADV 426
QY 279 P-QHVFGVADWNKPKMNGHDFADLKHFGNGGKYGKVMRYSDRDADSNYAFAGS----- 332
Db 427 PYSRVNPD-DPIDTKSFSIAYNLEHFRFSDNWTLRNSFRYLQQDLFTQATLAGSLNETT 485
QY 333 KLGKMTAGRPCGNTADDKACAVGLGTEIKOKALAFDASYSRPRNGTANEFVIGADYN 392
Db 486 GILTRTYAQR-----EYKSDYSIQTNNVVGKFTTGSIKKHTLLAGVDEN 528
QY 393 R-----FRST-----NEQGRITLYARGGLALNEF 416
Db 529 RGLLDDLVRGRTTLNIPNVPVGPRTDFTLPPATPFKNETRLGLYLOQQTALNN- 587
QY 417 RSIPQVDLIANARKVGRVGSHTVATENLDFEIGYKSTFHPADGLSLIGGRGLHYKIES 476
Db 588 ----QFTVLGRLYDVTDFKDTDESKVD-----SAWSPVIGL----- 622
QY 477 GEGKTLHKASKTKFTGYAGAVDYDLNNDNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQFE 536
Db 623 -----VYKPVENLSIYTSYRSFVPSFSDANGDFLOPERGAGYE 662
QY 537 VGYKSYMDRLNARVSFYRMKDKNAAPLNPNKKTRVYALGKRVMEGVETEISGAVTP 596
Db 663 VGKAEQLQGNLFATLAYFDITKQNVST-ADPDVLGAS-VATGEQKRSIGIELSAVGTIAP 720
QY 597 KWOIHAGSYLHSHQIKTASNRDDGIFL-----LMPKHSANLWTTVYQV---TPELTIGG 647
Db 721 GWNIIAGYAYTDAEI-----SKDNTIAGVNRLEPGTPKHSANLWTTVEIQKGSILQGLGFL 775
QY 648 GVNAMSGITSSAGMHAGGYATFDMAAAVRF-----PKLKLQINADNIENRHYARV 699
Db 776 GVNIV-----GKRFQNFQNDVEVDSYFLTNAALFYRQNNWRVGLNFNIFDINY---I 825
QY 700 GGANTFN-----IPGSERTWTANLRSYF 722
Db 826 SSAATLTRTSIEPQGPFTVVGSIISLEF 853

RESULT 38
ID P72609 PRELIMINARY; PRT; 853 AA.
AC P72609;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Ferrichrome-iron receptor.
GN FHUA OR SLR1490.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiura M., Sasamoto S., Kimura T.,
```

QY 430 KGVRFVSHVATENLDEF-----GIYKSTFHADSLIGGRIGHYKIES 476
Dh 559 YGI-----ATRPBRDEFADVPEFGSOTDSIGVFVONQTLLEMLKLIVGGRFDINDGSS 612
QY 477 GEGKTLHKASKTKETGYAGAVYDLNNNSLYLSLSQLYPQNTLDDGKILKPRQNGQE 536
Dh 613 ASDERQQA----FSPRVGLVYQPIEPISLYTSFSFSFQDNGNRRDGSLLFVETQYE 668
QY 537 VGKGYSTYMDRLNARVSYFMKDKNAAA--PLNPNNKKTRYALGKRVWEGYETELISGAV 594

Wed Dec 18 08:35:28 2002

```

595 TPWKQIHAGSYLSHQSITASNSRDDGIFLLMPKHSANLMTTYQVTPETLTGGGVNMSG 654
725 LPGNLIIASVATDKAVTKDDNLQGNLLDGVFPNSASLWYEQ-----TCDFQG 776
655 ITSSAGMHAGG-----YATDAAARVFTPKLKLQINADNIENRHYA--- 697
777 LGFGLGLFVYGERQGLNNSFQLPSVVRTDASIFYR-RNNWRAGINNLNLFNVYIEDSG 835
698 -----RVGGANTFNIPGS 710
836 QRNRINPGEPTVRCGT 852

RESULT 40
Q8XVU8 PRELIMINARY; PRT; 689 AA.
AC Q8XVU8;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative ferrichrome-iron receptor protein.
GN RSC2729 OR RS00125.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnean S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT Genome sequence of the plant pathogen Ralstonia solanacearum.;
RL Nature 415:497-502(2002).;
DR EMBL; AL646071; CAD16436.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 689 AA; 74213 MW; 6E9B8AFCDBD092C5 CRC64;

Query Match 10.9%; Score 412; DB 16; Length 689;
Best Local Similarity 22.8%; Pred. No. 3.7e-16;
Matches 170; Conservative 127; Mismatches 330; Indels 120; Gaps 26;

QY 20 ALSSVFAQAQTADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGTXIPASLREIPQSVI 79
19 ALBEAVVRAQAGE-----RLRARATSATLTD-----TPLKDVPSQVGV 57
80 ITNQVQKDRNVDTFDQARKTPGLRVLNSDDGRSSVVA-RGY-----EYSEYNIDGLPAQ 133
58 VTRAALDGFATRLDALTALDWMVSGISRONLGGIADNFAGIRFAGDLNTGSDYLVNGFSA- 116
134 MQSINGTLPLNLFADRVVEVVRGSGLEFSSGEMGGIVNLVVRKPTKFAQFOGHAAGFTGTHK 193
117 -NRANSVPVDVTINIRALDVLKGSAAALYGRSDPGGIVNITRTPQKPSREITLAAGSHD 175
194 QYKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEKNNRHETFYAADWDINP----- 246
176 QYRLATELTGPLSGH-----FAYRLGLA-----AENNHGFRFSAASRYVIAPFTWLPT 225
247 -DTVLGAGLYQQRHL-APYNGLPADANNKLPSPQHVFGVADWN---KPKMNSHDVFAD 301
226 DDTVV--TYAEAAQLRAPFDGRGIVAINRQLGALPDSRFLGEPGDDTTVTKQSHQL--S 281
302 LKHVFNNGYGVKVMHRY-----DEDADSNVAFAGSKLGMKTPAGRPCNTADDKACAVG 356
282 VEHLQSGWKVDAGLSYRSRFLFKGSKSDASRLDDGRTLWQARED---YHANDLAGRVD 339
357 LGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNNRFRSTNEQGRITLYARGGLALNEF 416

340 LQGEV-----ATGPRHTLVAGADFYNFRYDPVLYRANPSASAPYADIF 384
417 RS1-----PQVDLIANARKGVGYSHYATENLDEFGYKSTFHPADGSLIGGRLGH 471
395 DPVYQPRPALRPPSSSTRESORG-----FGAFVQDQVTLTPQWKLLAGVRMDR 432
472 YKIESGECKTLHKAS--KTKFTGYAGAVYDNDNNSLSLSQLYTPOTNLDADGKLKLP 529
433 FLQHADNRLTGARVAQQQTAVSPRLGLVYQPTQLSLYANTARSFRPNTGVGAQNAFAP 492
530 RQGNFEVGYKGSYMDRLNARVSFYRMKDKNAA--PLNPNKKTRVAALGKRVMEVGE 587
493 ERGRGEVGAKELETADGRFGGTLDALYSIDKTNVLTDGSRDPVVFQRTAGAVRSR-----GVE 548
588 TEISGAVTPKQIHAQSYLSHQSITASNSRDDGIFLL-MPKHSANLMTTYQV-----TPE 642
549 LDVSGQLTPNLKVLGTYYATDARV-TADTVLPSGAPLSNIPHSASALGLYFEGAGSLGR 607
643 LTIGGG---VNAMSGITSSAGMHAGGYATFDMAAYRFTPKLKLQINADNIENRHYAARV 699
608 AGVGGGVVYGERAGNSTDNGFKLPAYATVRLNGVYQPTRALRLSLTNIQNLDFKRY--- 664
700 GGANTFN-----IPGSERTWTANLRYSF 722
665 --ASSYNELMWAFGAERQVTLAATYTF 689

RESULT 41
Q92YCI PRELIMINARY; PRT; 714 AA.
AC Q92YCI;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative ferrichrome-iron receptor.
GN RA0963 OR SMA1747.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSyma megaplasmid.;
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AE007282; AAK65621.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 714 AA; 78235 MW; 4DAB5F4862934A69 CRC64;

Query Match 10.9%; Score 411; DB 16; Length 714;
Best Local Similarity 24.6%; Pred. No. 4.5e-16;
Matches 187; Conservative 110; Mismatches 344; Indels 120; Gaps 30;

QY 12 MTAATVLAALSSSVFAAQ-----TADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGT 64
24 LLGCTAFFALTPLSPALSAQAVPEGDTTVLETVIHAAGGGSVLNTDEDSKIATETGA 83
65 -KIPASLREIPQSVIITNQVQKDRNVDTFDQARKTPGL--RVLSNDDGRSSVYARGYE 121
84 GKMPDILIIAPASVSVITSKEIEERAADTIEQVQVTVAGVTVDFYGSDDRDYDFDINGFT 143
122 YSEYNIDGLPAQMSQINGTLPLNLFADRVVEVVRGSGLEFSSGEMGGIVNLVVRKPTKAF 181
```

```

Db      144 PYRR-DGL-ALGHTPGRVREBPYAFERIEVLKGAASSSSFCGAAPGGSVNVTTPPSDR 201
QY      182 QGHAAGFGTHKQYKADVSGSLNSDGVGRVMAQTVGASPPRAKNNRHETFYAAD 241
Db      202 FGEVYTGGSFHSKEIGDFGDNLTADETL-----SYRLTKFQRSDAEIYISQ 250
QY      242 WDN-----PDVTLGAGYLYQORHLAPYNGLPADANNKL-PSLPQHVFGAD--W 288
Db      251 DDENFWGVGTWRPTDASTLTFIFD--HL-DKQGVPGSGHPIGTDPDRDQFGEPPDYF 307
QY      289 NKRRNSHDVFAFLKHVFGNGYGVKVMRYSDDRADSNVAFAGSKLGMKTPAGPGCNTA 348
Db      308 SEINRNSYSLFPD--HDFGNGLSFSSNARYSNLNDGFSAYIGS-----TP 351
QY      349 DDKACAVG--LGTEIKOKLALFADASYRPRFGTANEFVIGADYRFRSTNEQGRITL 405
Db      352 TDGSTVAGRYFPGNEKSTDPFVIDAHLVYASLNDVSRLLFGADYKIKESD----- 403
QY      406 YARGLALNEFRSIPQVDLIANARKGVRSHTVATENLDE--FGIYKSTFHPADGLSL 463
Db      404 -----SANFYAPAPSIDWEDPIYSGGPGAMAPYASTNNQQTNAIYLOQDLFPFDKLT 457
QY      464 IGGGR-----LGHYKIESGEGKTLHKASKTKFTGYAGAVDLDNNSLYSLSQLYTPQT 518
Db      458 SFGLRNDWLDLSETNLLAG--TRACGNHREFTTRIGASYKYTEELAPYISYAESAP-- 512
QY      519 NLDADEKLLKPROGNOFEVYKGSYMDRLNA--RVSFYMKOKN-----AAPLNPNNK 571
Db      513 --PAAS--DPTGKYQYEVGK--YRPDAPFAPMTASVYLTKGNITVFQOVYTLPTQVE 566
QY      572 KTRVYALGKRWMEGVETESGAVTPKQIHAQYSYLHSQIKTASNSRHDGIPLMPKQSA 631
Db      567 KYRH-----RGFELEAKAEVTNNISVLAAYSIDSKIEEPGANDGRMLRVKPNMA 618
QY      632 NIMTYQVTPBELTIGG--GVNA-----MSGITSSAGMHAGGYATPDMAAARFTPKLK 662
Db      619 SVNGYTLLEDGARGDWLFGELGARYTDAYYTSITNTSSBSA--VWFDAPFYKIOENTT 676
QY      683 LQINADINFN-RHYVARVGANTENIPGSERTWTANLRYSF 722
Db      677 POLNVNNLPDEKRVASKDSGAVYTN-PG-RSLIATILRQSW 714

RESULT 42
Q9EZ93 PRELIMINARY; PRT; 762 AA.
ID      09EZ93
AC      01-MAR-2001 (TREMBLrel. 16, Created)
DT      01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE      Ferichrome-iron receptor (Fragment).
OS      Zymomonas mobilis.
OC      Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC      Zymomonas.
OX      NCBI_TaxId=542;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ZM4.
RA      Shin I.S., Kang H.S.;
RT      "Zymomonas mobilis ZM4 fosmid clone 43E12 complete sequence.";
RL      Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF300471; AAG42412.1; -.
DR      HSSP; P06971; LOJO.
DR      InterPro; IPR001899; Gram_pos_anchor.
DR      InterPro; IPR000531; TonB_boxC.
DR      Pfam; PF00593; TonB_boxC; 1.
DR      PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KM      Receptor.
FT      NON TER
SQ      SEQUENCE 762 AA; 83046 MW; F489525A093B358E CRC64;
Query Match 10.9%; Score 410.5; DB 2; Length 762;

```

```

Best Local Similarity 25.2%; Pred. No. 5,2e-16;
Matches 197; Conservative 111; Mismatches 319; Indels 155; Gaps 39;

QY      32 DLEFVHIKGRSVNATITEKNGDYSSPAVTVGKIPASLREIPOSVLIINQYKDNVD 91
Db      41 EVKTPAEKKKDKSDVIVVATG-LSQASSTTKTHTP--IIIESQISIVSRDEIELASP 97
QY      92 TFDOLARKTPGLRV-LSNDDGR--SSVYARGV---EYSEVN-IDG--LPQWOSINGTL 142
Db      98 TINDALAYTGVQAEPBGIDSRVDEVSVRGFGAGCSNNNPFVDGLRLPGSGQWTR-TSF 156
QY      143 NLPAFDRVEVMRGPSSGIFDSSGEMGIVNLVRRKPTPAQGH--AAAGFTHK---QYK 196
Db      157 DPESLQOIIEVLKPGSGALYQVATPGGVNVLVTRPKTKSGGEFFLQAGY-TLIGMWQGG 215
QY      197 AEAADVGSLSLNSDGVGRVMAQTVGASPPRAEKNNRHET--FPAAD-WDINPDT---- 248
Db      216 ASGDVSEKLNKDGTLGRI-----VGLARYGDTQVNHVQTRYYVSPSLTWEISPDTKWTV 271
QY      249 -----VLGAGYLYQORHLAPYNGLPADANNKLPSLPQHVFGADWKNKF 291
Db      272 LAQYQRDQGGSTQFLPATGTLYASKGRHA-----NDANIEPK-----WDF 315
QY      292 KNSHDVFAFLKHVFGNGYGVKVMRYSDDRADSNVAFAGSKLGMKT---PAGRPGCNT 347
Db      316 NRQVMAAGSPEFHEHFNH---FTLRNNTRYTYLNTLYRATVLSGDTLQSCSAIAGCVA 371
QY      348 AD--DCAVAGLGEIKOKALAPADASYRPRFGTANEFVIGADYRFRSTNEQGRITL 405
Db      372 GETVNRRAVEGRG--RSQGVATDTQLEGHVDTGAVRHVWLFGTDVFTYDWTDRD---- 424
QY      406 YARGLALNEFRSIPQVDLIANARKGVRG-----YSHVATENLDEFGIY----- 450
Db      425 LVYSSLVL-----PMLNIFDPVRSAGYGAALSPOVYITSTISQDN---GIYPQDQK 474
QY      451 -----GKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKFTGYAGAVYDLN 501
Db      475 FKRLRVTVGGROQWADDTLNLRTGKR-----YLTSHNAPFTWRAGAVYLPD 520
QY      502 DNNSLYLSLSQLYTPQT--NLDADEKLLKPROGNOFEVYKGSYMDRLNAEVSRYRK 558
Db      521 NGLAPFSYAKSFQPOVSPSTSLDTPPKPTGGDYEAIGIRYQ-GKSIYITFGAYQILT 579
QY      559 DKVAAPLPNPNKKTRVYALGKRWEGV--ETREISGAVTPKQIHAQYSYLHSQIK-TS 615
Db      580 QRMMTTP-DPNGALCSTICLVOTEIEGVPVGCLEBRATLPMNMALIFGTNRDAVYOS 638
QY      616 NSRD-DGFL-LMPKHSANIMWTYQVT---DELRTGGVNAMS--GITSSAGMHAGGY 666
Db      639 NTASQIGNYLPQAPKMAASLFDIDRIRHGFPTGIGGGVTRGHSYGDTNNT-LSIPG 697
QY      667 ATPDMAAARFTPK-----LKLQINADNI FNRHYVARVGANTENIPGSERTWTANLRY 720
Db      698 TLFDSFLRYDFANRPFPRYKGLSLISIMRNIAKRYVATCTAASACYV-QGGRSLTLRLGY 756
QY      721 SF 722
Db      757 SW 758

RESULT 43
Q8XVB1 PRELIMINARY; PRT; 801 AA.
ID      Q8XVB1
AC      08XVB1;
DT      01-MAR-2002 (TREMBLrel. 20, Created)
DT      01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE      Putative HYDROXAMATE-type ferrisiderophore receptor signal peptide
DE      protein.
GN      RSC2920 OR RSO0173.
OS      Ralstonia solanacearum (Pseudomonas solanacearum).
OC      Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC      Ralstonia.

```

[illegible]


```

Db 478 -----ADKEFVD-----KALAKGIFNNALOR-----FNSLDSSE 510
QY 503 NNSLYLSLQLYTPQTN-----LDADGKLKPRQGNQFVYKGSYMDRINAR- 551
Db 511 NKKATNRRSYSTVPLHTHTDODRQMGIKLDTGTYYGLFGREHDFVGY--AYGDEKIRSEY 568
QY 552 VSFYRKDKNAAPLNPNNKTRYAALGKRWEGVETEI-----SGAV 594
Db 569 LEYERRHR-----VAPNTGATGTVYAGS--CGSRTVICLLMSGGI 609

RESULT 45
P72599 PRELIMINARY; PRT; 863 AA.
AC P72599;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Ferrichrome-iron receptor.
GN FHDA OR SL11409.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Aemitsu B., Nakamura Y.,
RA Miyajima N., Hirose M., Sugitani M., Saeomoto S., Kimura T.,
RA Horouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tadaka S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90899; BAA16599.1;
DR InterPro; IPR000531; TOMB_boxC.
DR Pfam; PF00593; TOMB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 863 AA; 95026 MW; F8190445F9063CE4 CRC64;

Query Match 10.6%; Score 402; DB 16; Length 863;
Best Local Similarity 22.2%; Pred. No. 2e-15;
Matches 171; Conservative 127; Mismatches 335; Indels 136; Gaps 31;

QY 9 RINMTA-ATVLA-----LSSVFAAQTLDELTVNHKQGRSYNAIVT-EKNGDYSSE- 58
Db 150 RIVTNGNGVPAQVIPSSENLISLTPRINTV--ESEEEIRIVATREBEAAVOEFFVFN 207
QY 59 -AVTVGTIKIPASIREIPQSVIITNQOVDRVVDTPDOLARKTPGRLVLSNDGRSSVYA 117
Db 208 TSVAGTDTLP--IMDTFSAQVSEVIRSQQAITLEDEVLTNVSSVTTCGTTGGRETIIFG 265
QY 118 -RGY--EYSEYNIDGLPAQMOSIN-----GTLPNLFAFDRYEVMRGPGLPDSGEMGI 169
Db 266 IRGFENQFS-----DTVPILRDGFRLYGGFGQITVSHLQOVEVLKGPSSILVGQIEPGV 321
QY 170 VNLVXRPTKAFQGHAAAGFGTHKQYKAADVSGSLNSDGSYRGRV--AQTYGASPPRAE 228
Db 322 INLSKSKPLNEPFAVEVOLGNGLVPRFRDISGGLNPSGNLRYRLNGVYSNEASFRDFN 381
QY 229 KNNHETFYAADMDINDPTVLGAGVLYQORHLAPYNGI-----PAD 270
Db 382 QPLERFAFAPIVTVAITLDTDLASLAVEYINDTNPADPGLSSFGDGAAPVPSRVINDPSD 441
QY 271 ANNTKLPSLPQHFVFCADNNKFKMNSHDVPAFLKHYFGNGGYGKVGMRYSDDRADSNYAF 330
Db 442 IVNK-----NFTSAGYN-----LEHRFNEWMKLRNAPRYMSYVDYVNV- 479
QY 331 GSKGKMTFAGRPQNT--ADDKACAVGLGTRIKOKALAFDASYSRPRLTANTANEFVIG 388
Db 480 ---IALPTVNGPVTVPFADODG-----QGSYSFTYNAGVGFSTGSVKHELLAG 527

```

```

QY 389 ADVNFRSTNEQRTLLVARGGLAN-----EFSRIPQVDLINARKVGYSHTVATENU 444
Db 528 IDYN-----WSESLTLTFG-GPRTSIVFPDPDVAIPKN-----RDDLPLFGDTFSSN- 576
QY 445 DEFGIYKSTFHPADGLSLIGGRLGHYKIESGEKTLHKASKTKFTGYA-----GAVY 498
Db 577 -RLGIYLQOVSLLENLILVAGLRVDTITQNTNNLQTDENQGNTOQTDSAVTPRIGLLY 635
QY 499 DLNDNNSLYLSLQLYTPQTNLDADGKLKPRQGNQFVYKGSYMDRINARVSFRNK 558
Db 636 RPIPEISFFSNYSQSFTPNSGIDISGNPLPERGEGFEGVAEALFEQQLTLTYFNIS 695
QY 559 DKNAAA--PLNPNKKTTRYAALGKRWEGVETEISGAVTPKQOIHAGSYLHSQI--KTA 614
Db 696 KNNVAVSDPVNPLFST-----IGTQOSQSELDIYGEIILPGMKIIGNYSINAKVTEPD 751
QY 615 SNSRDDGIFLLMPKHSANIMWTYQVTPBELTTIGGVNAMSGITSSAGM----- 662
Db 752 PNFVDNRLFGI--PYNNANLMTTYEIQS-----GALQGLGFGIGFNVVGDPRFDLANT 802
QY 663 --AGGATDTMAAAYRFTPKLQLQINADNIFNRHYIYARVGANTRNIPG 709
Db 803 YTVGDYIIIGNAAIFYQ-RDKYRVALLNRNFTNANYVRAVSGNQGTIEPG 850

```

Search completed: December 18, 2002, 06:44:53
 Job time : 62.0531 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 05:44:19 ; Search time 44.9653 Seconds
(without alignments)
2139.581 Million cell updates/sec

Title: US-09-889-267-2

Sequence: 1 MGQFM~~SV~~FRINMTAATVLA.....NTFNIPGSE~~RT~~WTANLRYSF 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : **A_Geneseq_101002.***

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/Aa2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query Match	Length	ID	Description
1	3776	100.0	722	21	AAB07697
2	3592	95.1	691	21	AAB07698
3	3244	85.9	708	21	AAV75566
4	821.5	21.8	815	22	AAU33610
5	784	20.8	725	20	AAV38631
6	784	20.8	725	20	AAV38632
7	778	20.6	725	20	AAV38633
8	765	20.3	154	21	AAV75565
9	737.5	19.5	729	22	AAV34539
10	676.5	17.9	623	20	AAV38633

11	400.5	10.6	783	21	AAV75553	Neisseria meningit
12	397.5	10.5	783	21	AAV75554	Neisseria meningit
13	375.5	9.9	702	22	AAAB88923	Neisseria meningit
14	374.5	9.9	702	22	AAAB88921	Neisseria meningit
15	371.9	9.8	663	22	ABBS2462	Escherichia coli p
16	370	9.8	703	22	AAAB88926	Neisseria meningit
17	370	9.8	703	21	AAV75568	Neisseria meningit
18	370	9.8	703	21	AAAB88916	Neisseria meningit
19	368	9.7	703	22	AAV75567	Neisseria meningit
20	368	9.7	703	22	AAAB88927	Neisseria meningit
21	367.5	9.7	702	22	AAAB88920	Neisseria meningit
22	366	9.7	703	21	AAAB18720	A Neisseria mening
23	364.5	9.7	702	22	AAAB88922	Neisseria meningit
24	363	9.6	703	22	AAAB88919	Neisseria meningit
25	316.5	8.4	697	22	AAAB88917	Neisseria meningit
26	316.5	8.4	697	22	AAAB88925	Neisseria meningit
27	313.5	8.3	721	22	ABBS2269	Escherichia coli p
28	307.5	8.1	700	22	AAAB88918	Neisseria meningit
29	294.5	7.8	696	17	AAAT5356	Adhsin, Escherich
30	273.5	7.2	725	22	AAAT72924	E. coli ironEc ext
31	241.5	6.4	697	22	AAAB88924	Neisseria meningit
32	237.5	6.3	811	20	AAU33589	Pseudomonas aerugi
33	235.5	6.2	811	20	AAV788830	Neisseria meningit
34	228	6.0	660	22	ABBS2269	Escherichia coli p
35	222.5	5.9	682	22	ABBS22836	Escherichia coli p
36	220	5.8	785	21	AAAT80056	TonB dependent rec
37	213	5.6	187	21	AAAT75536	Neisseria meningit
38	205	5.4	188	21	AAAT75535	Neisseria meningit
39	206	5.4	774	22	AAU729336	Novel mar regulate
40	195	5.2	454	22	ABG17860	Novel human diagno
41	195	5.2	134	21	AAAB40909	Human ORFX ORP673
42	195	5.2	134	22	ABPB60766	Human ORFX ORP673
43	195	5.2	340	23	ABBS2269	Novel human diagno
44	193.5	5.1	833	20	AAV344944	Porphyromonas ging
45	193.5	5.1	876	20	AAV34358	Porphyromonas ging

ALIGNMENTS

```

RESULT 1
AAB07697
ID AAB07697 standard; Protein; 722 AA.
XX
AC AAB07697;
XX
DT 07-NOV-2000 (first entry)
XX
DE A Neisseria meningitidis BASB053 polypeptide.
XX
KW BASB053; Neisseria meningitidis infection; vaccine.
XX
OS Neisseria meningitidis.
XX
PN WO200042193-A1.
XX
PD 20-JUL-2000.
XX
PF 10-JAN-2000; 2000WO-EP00137.
XX
PR 15-JAN-1999; 99GB-0000959.
PR 28-JAN-1999; 99GB-0001903.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
Ruelle J;
XX
PI
XX
WP1: 2000-476062/41.
DR
DR N-PSDB; AAA59216.
XX
FT New Neisseria meningitidis polypeptide useful for diagnosis of
FT Neisseria infection and for development of vaccines against such
FT infection -
FT

```

```
XX PS Claim 3; Page 55-56; 92pp; English.
XX CC The present sequence represents a Neisseria meningitidis BASB053
CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific
CC for BASB053 may be identified in a biological sample in order to
CC diagnose a Neisseria meningitidis infection in an animal. The BASB053
CC polypeptides and polynucleotides may be used as vaccines, for
CC generating an immune response in an animal. A composition comprising
CC at least one antibody immunospecific for BASB053 may be used to
CC treat humans infected with Neisseria meningitidis.
XX SQ Sequence 722 AA;
Query Match 100.0%; Score 3776; DB 21; Length 722;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQQFMSVFRINNTAATVLAALSSVFAAQTADLETWHIKQORSYNAIVTEKNGDYSSFAV 60
Db 1 MQQFMSVFRINNTAATVLAALSSVFAAQTADLETWHIKQORSYNAIVTEKNGDYSSFAV 60
Qy 61 TVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDOLARKTPGLRLVLSNDDGRSSVYARGY 120
Db 61 TVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDOLARKTPGLRLVLSNDDGRSSVYARGY 120
Qy 121 EYSEYNIDGLPAQMOSINGTLPNLFAPDRVEVMRGPGLFDSSGEMGGIVNLVRKRPTKA 180
Db 121 EYSEYNIDGLPAQMOSINGTLPNLFAPDRVEVMRGPGLFDSSGEMGGIVNLVRKRPTKA 180
Qy 181 FQCHAAAGFTGHQYKAEADVSGSLSDGSRVGRVMAQTVGASPRPAEKNNRHETFYAAA 240
Db 181 FQCHAAAGFTGHQYKAEADVSGSLSDGSRVGRVMAQTVGASPRPAEKNNRHETFYAAA 240
Qy 241 DWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLPSPLOHVFVCGADNKFQNNSHDVA 300
Db 241 DWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLPSPLOHVFVCGADNKFQNNSHDVA 300
Qy 301 DLKHVFGNGGYGKVMRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTE 360
Db 301 DLKHVFGNGGYGKVMRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTE 360
Qy 361 IKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEQGRITLYARGGLALNEFRSIP 420
Db 361 IKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEQGRITLYARGGLALNEFRSIP 420
Qy 421 QVDLIANARKGVGYSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGK 480
Db 421 QVDLIANARKGVGYSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGK 480
Qy 481 TLHKASKTKFTGYAGAVYDLNDNNSLSLSQLYTPQTNLDADGKLLKPRQGNQFEVGKY 540
Db 481 TLHKASKTKFTGYAGAVYDLNDNNSLSLSQLYTPQTNLDADGKLLKPRQGNQFEVGKY 540
Qy 541 GSYMDRLNARVSFRMKDKNAAPLNPNNKTRTAALGKRWMEGVETEISGAVTPKQKI 600
Db 541 GSYMDRLNARVSFRMKDKNAAPLNPNNKTRTAALGKRWMEGVETEISGAVTPKQKI 600
Qy 601 HAGSYLSHSQIKTASNSRDDGIPLMPKHSANLWTTYQVTPELTIGGGVNMAGTSSAG 660
Db 601 HAGSYLSHSQIKTASNSRDDGIPLMPKHSANLWTTYQVTPELTIGGGVNMAGTSSAG 660
Qy 661 MHAGGYATFDMAAAYRFTPKLQLINADNIFNRHYARVGGANTFNI PGSERTWTANLRY 720
Db 661 MHAGGYATFDMAAAYRFTPKLQLINADNIFNRHYARVGGANTFNI PGSERTWTANLRY 720
Qy 721 SF 722
Db 721 SF 722
RESULT 2
AAB07698
```

```

ID XX AAB07698 standard; Protein; 691 AA.
XX AC AAB07698;
XX DT 07-NOV-2000 (first entry)
XX DE A Neisseria meningitidis BASB053 polypeptide.
XX KW BASB053; Neisseria meningitidis infection; vaccine.
XX OS Neisseria meningitidis.
XX PN WO2000042193-A1.
XX PD 20-JUL-2000.
XX PF 10-JAN-2000; 2000WO-EP00137.
XX PR 15-JAN-1999; 99GB-0000959.
XX PR 28-JAN-1999; 99GB-0001903.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Ruelle J;
XX DR WPI; 2000-476062/41.
XX DR N-PSDB; AAA59217.
XX PT New Neisseria meningitidis polypeptide useful for diagnosis of
XX PT Neisseria infection and for development of vaccines against such
XX PS infection -
XX PS Claim 3; Page 56; 92pp; English.
XX SQ Sequence 691 AA;
Query Match 95.1%; Score 3592; DB 21; Length 691;
Best Local Similarity 99.7%; Pred. No. 3.8e-318;
Matches 684; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 37 HIKQORSYNAIVTEKNGDYSSFAVTVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDQL 96
Db 6 HIKQORSYNAIVTEKNGDYSSFAVTVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDQL 65
Qy 97 ARKTPGLRVLNDDGRSSVYARGVEYSEYNIDGLPAQMOSINGTLPNLFAPDRVEMRGP 156
Db 66 ARKTPGLRVLNDDGRSSVYARGVEYSEYNIDGLPAQMOSINGTLPNLFAPDRVEMRGP 125
Qy 157 SGLFDDSSGEMGGIVNLVRKRPTKAFQGHAAAGFTGHQYKAEADVSGSLSDGSRVGRVM 216
Db 126 SGLFDDSSGEMGGIVNLVRKRPTKAFQGHAAAGFTGHQYKAEADVSGSLSDGSRVGRVM 185
Qy 217 AQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLP 276
Db 186 AQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLP 245
Qy 277 SLPOHVFVGADWNKFKMNSHDVFDLKHYPGNGGYGKVMRYSDRDADSNYAFAGSKLGM 336
Db 246 SLPOHVFVGADWNKFKMNSHDVFDLKHYPGNGGYGKVMRYSDRDADSNYAFAGSKLGM 305
Qy 337 KTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRS 396
Db 306 KTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRS 365
```

QY 397 TNEGGTTLYARGGLANEFRRSIPQVDLIANARKVGRGSHYVATENDPFGYKSTPH 456
 Db 366 TNEGGTTLYARGGLANEFRRSIPQVDLIANARKVGRGSHYVATENDPFGYKSTPH 425
 QY 457 PADGSLIGGRLGHYKISSEGGKTLHKASKTFGTGAGAVYDLDNNNSLYLSLSQLYTP 516
 Db 426 PADGSLIGGRLGHYKISSEGGKTLHKASKTFGTGAGAVYDLDNNNSLYLSLSQLYTP 485
 QY 517 QTNLDADGKLLKPRQNGQEVGKGSYMDRLNARVSFYRMKDKNAAAPLNNKKTTRYA 576
 Db 486 QTNLDADGKLLKPRQNGQEVGKGSYMDRLNARVSFYRMKDKNAAAPLNNKKTTRYA 545
 QY 577 ALGKRVMEGVEFETISGAVTPKMOIHAGYSLHSQITKASNSRDDGIFLLMPKGSANLMTT 636
 Db 546 ALGKRVMEGVEFETISGAVTPKMOIHAGYSLHSQITKASNSRDDGIFLLMPKGSANLMTT 605
 QY 637 YQVTPPELTIGGGVNAMSGITSSAGMAGGATPDMAAVRFPPKLTQINADNIFNRHY 696
 Db 606 YQVTPPELTIGGGVNAMSGITSSAGMAGGATPDMAAVRFPPKLTQINADNIFNRHY 665
 QY 697 ARVGGANTFNIPGSERTWTANLRYSF 722
 Db 666 ARVGGANTFNIPGSERTWTANLRYSF 691

RESULT 3
 AAY75566
 ID AAY75566 standard; Protein; 708 AA.
 XX AAY75566;

XX 21-MAR-2000 (first entry)
 DE Neisseria meningitidis ORF 760 protein sequence SEQ ID NO:2606.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.

XX Neisseria meningitidis.

XX WO9597280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0098994.

XX 09-OCT-1998; 98US-0103794.

XX 09-OCT-1998; 98US-0103794.

XX 09-OCT-1998; 98US-0103796.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masigman V, Mora M;
 PI Petersen J, Pizzo M, Rappoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelein H, Venter JC;

XX MPI: 2000-062150/05.
 DR N-PSDB; AA254328.

CC Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX Claim 2; Page 1235; 1453bp; English.
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254615 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 708 AA;

Query Match 85.9%; Score 3244; DB 21; Length 708;

Best Local Similarity 86.8%; Pred. No. 2,5e-286;
 Matches 627; Conservative 33; Mismatches 48; Indels 14; Gaps 3;

QY 1 MGFMSVFRINNTAATVLAALSSVFAAQTADLETVHIKGRSYNAIVTERKNDYSSFAV 60
 Db 1 MGFMSVFRINNTAATVLAALSSVFAAQTADLETVHIKGRSYNAIVTERKNDYSSFAV 60
 QY 61 TVGTKIPASLRREIPQSVSITITQVADRVDITFDLARTPLRLVLSNDGSSVYARGY 120
 Db 61 TVGTKIPASLRREIPQSVSITITQVADRVDITFDLARTPLRLVLSNDGSSVYARGY 120
 QY 121 EYSEYVNDGLPAWOSINGTLPLNLFAPDRVEVMRGPGLFDSGEMGIVNLVRRKPTA 180
 Db 121 EYSEYVNDGLPAWOSINGTLPLNLFAPDRVEVMRGPGLFDSGEMGIVNLVRRKPTA 180
 QY 181 FQGHAAAGFTHKQYKAADVSGSINSDSVGRVMAQTGASPRPAEKNNHEFTYAAA 240
 Db 181 FQGHAAAGFTHKQYKAADVSGSINSDSVGRVMAQTGASPRPAEKNNHEFTYAAA 240
 QY 241 DWDINPDTLGGHYLYQOCHLAPYNGLPADANKLPSPQHVFVGDMMKFFKMSHDVA 300
 Db 241 DWDINPDTLGGHYLYQOCHLAPYNGLPADANKLPSPQHVFVGDMMKFFKMSHDVA 300
 QY 301 DLKHYFNGGKGVKGRYSRDRDASNYAPAGSKLGWKTAPGRPGCNTADKCAVGLGTE 360
 Db 301 DLKHYFNGGKGVKGRYSRDRDASNYAPAGSKLGWKTAPGRPGCNTADKCAVGLGTE 360
 QY 361 IKQALAPDASYSRPRLGNTANEFVIGADYVRFSTNEQSTTLYARGGLANEFRRSIP 420
 Db 349 IKQAFADVASYSRPALGNTANEFVIGADYVRLNSTNEQGSTL--SKSVALDGRFALP 406
 QY 421 QVDLIANARKVGRGSHYVATENDPFGYKSTHPADGLSIGGRLGHYKISSEGGK 480
 Db 407 YNGIIONARAGKGFHNSYTEENLDETGLYAKTVRPLRGLSLIAGRGVGHKIESGDGK 466
 QY 481 TLHKASKTFGTGAGAVYDLDNNNSLYLSLSQLYTPQTNLDADGKLLKPRQNGQEVGK 540
 Db 467 TLHKASKTFGTGAGAVYDLDNNNSLYLSLSQLYTPQTNLDADGKLLKPRQNGQEVGK 526
 QY 541 GSYMDRLNARVSFYRMKDKNAAAPLNNKKTTRYAALGKRVMEGVEFETISGAVTPKMOI 600
 Db 527 GSYMDRLNARVSFYRMKDKNAAAPLDSNNKKTTRYAALGKRVMEGVEFETISGAVTPKMOI 586
 QY 601 HAGYSLHSQITKASNSRDDGIFLLMPKGSANLMTTYQVTPPELTIGGGVNAMSGITSSAG 660
 Db 587 HAGYSLHSQITKASNSRDDGIFLLMPKGSANLMTTYQVTPPELTIGGGVNAMSGITSSAG 646
 QY 661 MHAGGYATPDMAAAYRFPKLTQINADNIFNRHYARVGGANTFNIPGSERTWTANLRY 720
 Db 647 IHAGGYATPDMAAAYRFPKLTQINADNIFNRHYARVGGANTFNIPGSERTWTANLRY 706
 QY 721 SF 722
 Db 707 SF 708

RESULT 4
 AAU33610

AAU33610 standard; Protein; 815 AA.	
AAU33610;	
14-FEB-2002 (first entry)	
Pseudomonas aeruginosa cellular proliferation protein #54.	
Antisense; prokaryotic cellular proliferation protein;	
antibiotic; antibacterial; drug design.	
Pseudomonas aeruginosa.	
WO200170955-A2.	
27-SEP-2001.	
21-MAR-2001; 2001WO-US09180.	
21-MAR-2000; 2000US-191078P.	
23-MAY-2000; 2000US-206848P.	
26-MAY-2000; 2000US-207727P.	
23-OCT-2000; 2000US-242578P.	
27-NOV-2000; 2000US-253625P.	
22-DEC-2000; 2000US-257931P.	
16-FEB-2001; 2001US-269308P.	
(ELIT-) ELITRA PHARM INC.	
Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;	
Yamamoto RT, Xu HH;	
WPI; 2001-611495/70.	
N-PSDB; AAS51469.	
New polynucleotides for the identification and development of	
antibiotics, comprise sequences of antisense nucleic acids -	
Example 3; Seq ID No 5106; 511pp; English.	
The invention relates to antisense inhibitors of genes essential to	
prokaryotic cellular proliferation, their use in identifying the	
genes, themselves and in the discovery of novel antibiotics, the essential	
genes, themselves and the encoded proteins. The prokaryotes used are	
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella	
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The	
invention is also useful for the identification of potential new targets	
for antibiotic development. The antisense nucleic acids can also be used	
to identify proteins used in proliferation, to express these proteins,	
and to obtain antibodies capable of binding to the expressed proteins.	
The proteins can be used to screen compounds in rational drug discovery	
programmes. The antisense nucleic acid sequence is also useful to screen	
for homologous nucleic acids which are required for cell proliferation in	
a wide variety of organisms. The present sequence represents an	
essential prokaryotic cellular proliferation protein.	
Note: The sequence data for this patent did not form part	
of the printed specification, but was obtained in electronic	
format directly from WIPO at	
ftp.wipo.int/pub/published_pct_sequences.	
Sequence 815 AA;	
Query Match 21.8%; Score 821.5; DB 22; Length 815;	
Best Local Similarity 29.0%; Pred. No. 2e-65;	
Matches 217; Conservative 145; Mismatches 312; Indels 75; Gaps 24	
8 FRINMTAATVLAALSSSVFAAQTADLETVHKGQRSYNAIVTEKNGDYSSFAVTGVTKIP 67	
108 FQGNATISVAEADSSV-----DLGATMITSNQL--GTITEDSGSVTPGTATATRLV 159	
68 ASUREITPQSVSIITNOQVKRNVDTFDQLAKTPEGLRVLSNDDGRSSVYARGVEYSYNI 127	
160 LTPRETPOSITVTRQMDDFGLNIDDDVMRHTPGITVSAYDTRNNYVARGFSINNFQY 219	

Qy	128	DGLPAQMOSI-----NGTLPNLFPADRVVWRGSPGLFDSSGEMGGIVNLVRKRPPTKAPQG	103
Db	220	DGIPSTARNVGYSGAGNTLSDMAIYDRVVELVKGATGLLTGAGSLGATINLIRKKPTHEFKG	279
Qy	184	HAAGAFGTHQYKAEADVSGSLNSDGSGVRGVMAQTVGASPRPAEKNNRHETFYAAADWD	243
Db	280	HVELGAGSWNYKSELDSGLPTESGNVRGRVAAVQDKHISFMDHYERKTSVYVIGILEFD	339
Qy	244	INPDTVLGAGLYLQORHL--APYNG-LPA-DANNKLPSLPQHVFVGADWNKFKQNSHDVF	299
Db	340	LNPDMLTVGADYQDNPKKSGSGSGFPFLPDSQGNRNDVSRSPNNKAGKSSWQYTRTFV	399
Qy	300	ADLKHYPGNGGYKGVKWRVSDRDAISNYPAGSKLGMKTPAGRPAGCNATDDKACAVGLGT	359
Db	400	ANLEHNPANGWGVQVL---DHKINGYHAPLGAIMG-DWPA-----PDNSAKIVAQKY	448
Qy	360	EIKOKALAFDASVSRPRLGNTANEFVIG--ADYNFRSTNEQGRITLYARG-GLALNEF	416
Db	449	TGETKSNLSLDIYTGPPQFLGREHELVLVGTSASFHW-----EGKSYWNLRNVDNTDDF	503
Qy	417	RSIPQVDLIAANARKGVRGYSHTVATENLDEFGYIGKSTFHPADGLSLIGGRLGHYKIES	476
Db	504	-----INWGDIGKPDWGTSPQYIDDKTRGLQSYMTAREFNVTDNLFLGGRVVDYRV-T	557
Qy	477	GEKTLHKASKTFETGYAGAVYDLNDDNSLYLSLSOLYTPQTV--LDADGKLLKPRQNGQ	534
Db	558	GLNPTITRESG--RFPIYVGAVYDLNDDNTSYSYASTYDIFMPQDSWYRDSNNKLLPEDEGQ	615
Qy	535	FEVGYKGSYMDRLNARVSFYRMKDKNAAAP-----LNPNNKTRVAAALG-KRVMSGVET	588
Db	616	YEIGIKGEYLDGLRLNTSLAYFEIHEENRAEEDALYNSKPTNPATITAYYKGIKAKTKGYEA	675
Qy	589	EISGAVTPKQIHHAGYSYLHSQIKTASNGRDDGIFLLMPKHSANLWTTYQ---VTPELTI	645
Db	676	EISGELAPGHQVQAGYT--HKIIRDDSGKK--VSTWEPQDQLSLVTSYKFKGALDKLTV	730
Qy	646	GGGV-----NAMSGITSSAGMHAGGYATFDMAAARFTPEKLLQINADNIFNR	693
Db	731	GGGARWGKSWQWYNNPRSRWEK---FQGEDYWLVDLMARYQITDKLSASVNVNNVFVK	787
Qy	694	HYTARVGGANTFNIPGSRERTWTANLRYSF 722	
Db	788	TYTNTIGFYTASY-GDPRNLMESTRWDP 815	
RESULT 5			
AAV38831			
ID	AAV38831 standard; Protein; 725 AA.		
XX	AAV38831;		
AC			
XX			
XX	08-OCT-1999 (first entry)		
DT			
XX	Neisseria meningitidis antigen encoded by ORF23.		
DE			
XX	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		
KW	treatment; Neisseria infection; meningitis; septicemia; gonorrhea.		
XX			
OS	Neisseria meningitidis.		
XX			
PN	WO9924578-A2.		
XX			
PD	20-MAY-1999.		
XX			
PF	09-OCT-1998; 98WO-IB01665.		
XX			
PR	01-SEP-1998; 98GB-0019016.		
PR	06-NOV-1997; 97GB-0023516.		
PR	14-NOV-1997; 97GB-0024190.		
PR	18-NOV-1997; 97GB-0024386.		
PR	27-NOV-1997; 97GB-0025158.		
PR	10-DEC-1997; 97GB-0026147.		

```

PR 14-JAN-1998; 98GB-0000759.
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Maignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI, 1999-327407/27.
XX N-PSDB; AA212259.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 379-380; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX
XX Sequence 725 AA.
XX
XX Query Match 20.8%; Score 784; DB 20; Length 725;
XX Best Local Similarity 29.4%; Pred. No. 4.5e-62;
XX Matches 228; Conservative 118; Mismatches 320; Indels 110; Gaps 21;

5 MSVFRINMTAATVL-----AALSSVFAAQTADLETVHIKQSSYNAIVTEKNGDYSS 57
1 MRFKXSLIFALLPVYAQADVSDPKPQESTELPITTVADRT-----ASSNDGYTV 55
58 FAVTVGTKIPASLREIPQSVSIITNOQVDRVDFDQ-LARKTPGLRYLSND--GRSS 114
56 SGTHTPPLGLPMTLREIPQSVSVITTSQMRDQNKITLDRLLQATGTSKQIYGSDBRAGVY 115
115 VYARGYEVSEYNIDGLPAQWQISNGTLPNLFAFDRVEVMKPSGLPDSGEMGIVNLVR 174
116 LPARGSRINAYQNGIPVADALADTGNANTAAVERVEVVRGVAGLLDGTGEPAATNLVR 175
175 KKPRTAFQGHAAAGFTTHQYKAEADVSGSLNSDGVKRWQAQTVGASBPRAEKNRHE 234
176 KRLTRKPLFEVRAEAGNRKHFGILDADVSGSLNTEGTLRGSLVTFPCRQDSWRRERSRA 235
235 TTYAADMINDINPTVVGAGVLYXQORHL---APYNGLPADANNLPSLPQHVFGAQMNF 291
236 ELVGIIEYIAQOTRYHAGMDYQAKETADAPLSIAYVDSGYATAPGPRDNPATMANS 295
292 KMSHDVFPADLKHYFNGGVYKVKMRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADK 351
296 RRRALNLPAFGEIHRFN-----QDWKLAEDYTRSR--PRQPGVAGVLSIDHN 342
352 ACAVGL---GTBKOKALAFDASYSRPFRLGNTANFVIGADYNRRSTNEGRTTLVAR 408
343 TATATDIPYWHADPRTHSASVSLICKYRLFGRHDLGAGINQYKYS--NKYGE----- 395
409 GGLALNEFSPISQVLDLIANARKGVRSHTVA-----TENLDERGI-----YKSTP 455
396 -----RSI-----IPNAIPNAYEFSRTGATPPAPASPAQTIPQYGTTRQIGCYATRR 442
456 HPADGLSLIGGRLGHYKLESGEKGT--LHKASKTFGTGAGAVYLDNNNSLYLSLOU 513
443 RAADNLSTLIGGRTYTRRGYSYDSRTQGMTYVANSRFTPTGTGVPFLTGNLSYGSYSL 502
514 YTPQTLVDADGLLKRPQGNQFVYKGSYMDRLNARVSFTYMKQKN--AAAPLNPNNK 571
503 FVPQSQKDEHGSYLKPEVTGNLLEAGIKGEWLBGRILNASSAAYVPARKNLTATAGRDPSC- 561
572 KTRVALGKRVNVEGTEISGAVTPKMOIHAGSYLHSQIKTASNSRDGDIPLMFKHS 631
562 NITYRAANAQKTHGWEIVGGRTITPEMOIOAGISQSKTRDQDSRLNPDSV---PERSE 617

```

```

QY 632 NIMTYQVPEL-----TTCGV-----NANSGITSSAGMHAQVATPDA 671
DB 618 KLFAYVHFAPEAPSGWITGAGVRSQSETHDPATLRIPAPAKARAADSROKAYAVADI 677
QY 672 MAAYFTPELKIQIANDINFNHRYARVQANTFNP-----GSRRTWANLRYSF 722
DB 678 MARYRNPRAELSLVNDLNFKNHYRTQ-----PDRHSYGALRTVNAFTYR 724

RESULT 6
AAY38834
ID AAY38834 standard; Protein, 725 AA.
XX
XX AAY38834;
XX
XX 08-OCT-1999 (first entry)
XX
XX Neisseria gonorrhoeae antigenic protein encoded by ORF23.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
XX Neisseria gonorrhoeae.
XX
XX WO924578-A2.
XX
XX 20-MAY-1999.
XX
XX 09-OCT-1998; 98MO-IB01665.
XX
XX 01-SEP-1998; 98GB-0019016.
XX 06-NOV-1997; 97GB-0023516.
XX 14-NOV-1997; 97GB-0024190.
XX 18-NOV-1997; 97GB-0024386.
XX 27-NOV-1997; 97GB-0025158.
XX 10-DEC-1997; 97GB-0026147.
XX 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Maignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI, 1999-327407/27.
XX P-PSDB; AAY38834.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 384; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX
XX Sequence 725 AA;
XX
XX Query Match 20.8%; Score 784; DB 20; Length 725;
XX Best Local Similarity 29.6%; Pred. No. 4.5e-62;
XX Matches 230; Conservative 115; Mismatches 321; Indels 110; Gaps 21;

5 MSVFRINMTAATVL-----AALSSVFAAQTADLETVHIKQSSYNAIVTEKNGDYSS 57
1 MRFKXSLIFALLPVYAQADVSDPKPQESTELPITTVADRT-----ASSNDGYTV 55
58 FAVTVGTKIPASLREIPQSVSIITNOQVDRVDFDQ-LARKTPGLRYLSND--GRSS 114
56 SGTHTPPLGLPMTLREIPQSVSVITTSQMRDQNKITLDRLLQATGTSKQIYGSDBRAGVY 115

```

XX	(CHIR-) CHIRON SPA.
PA	
XX	
PI	Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX	
DR	WPI; 1999-327407/27.
DR	N-PSDB; AAZ12260.
XX	
PT	Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT	diagnosis, treatment and prevention of infection
XX	
PS	Claim 4; Page 381; 524pp; English.
XX	
CC	Amino acid sequences AA Y38499-Y38944 represent Neisseria meningitidis
CC	and N. gonorrhoeae antigenic proteins. They are encoded by open
CC	reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC	their fragments, their nucleic acids and antibodies are used for
CC	diagnosis, prevention (as vaccines) or treatment of Neisseria
CC	infections, such as meningitis, septicemia and gonorrhea. Both
CC	organisms are closely related. Fragments of the nucleic acids
CC	are useful as hybridisation probes and antisense reagents.
XX	
SQ	Sequence 725 AA;
	Query Match 20.6%; Score 778; DB 20; Length 725;
	Best Local Similarity 29.7%; Pred. No. 1.6e-61;
	Matches 231; Conservative 117; Mismatches 317; Indels 112; Gaps
Qy	5 MSVPRINMTATVL-----AALSSVFPAQTADLETVHIKQORSYNAIVTEKNGDYSS 57
Db	: :
	1 MTRPKYSLLFAALLPVYAQADVSDPKPQEISTELPTITVTADRT-----ASSNDGYTV 55
Qy	58 FAVTVGTKIPASLREIPQSIIITNOQVKDKNVDTFDQ-LARKTIPGLRVLSNDD--GRSS 114
Db	: :
	56 SGTHTPCLGFWTLREIPQSIVSVITSQQMRDQNKALDRALLQATGTSRQIYGSDRAGNY 115
Qy	115 VIARGYSESYINDIGLPAAQMOSINGTLPNLFAFDREVNRGPSLGSPDGSGEMGIIVNLR 174
Db	: :
	116 LPARGSIANYQINGIPVADALADTGNTANTAAVERVEVVGVAGLLDGTGEPSATVNLR 175
Qy	175 KRPTKAFQGHAAAGFTHKQYKAEADVSGSLNSDGSVGRGVMAQ-TVGASPPRAEKNNH 233
Db	: :
	176 KRPTRKPLFEVRAEAGNRKHFGVGADVSGSLNAEGTLRGRLVSTFGRGSWRQRER-SRD 234
Qy	234 ETFFAAAADWDINPDTVLGACLYQOHRHL---APVNGLPADANNKLPSLPQHVPVGDANK 290
Db	: :
	235 AELYGLEIYDIAQPTRVHAGWDYQAQKETHADAPLSVAVDISOQYATAFGPKONPATNWAN 294
Qy	291 FQMSHDVDFADLKHYFGNGGGYKGVGMRYSDRDADSNIYAFAGSKLGMKTAPGRGCNTADD 350
Db	: :
	295 SRHRALNLFAGIEHREFN-----QDWKLKAEYDVYTRS- --FRQPVGVAGVLSIDH 341
Qy	351 KACAVGL---GTEIKQALAFDASYSRPFRLGNATANEFVIGADYNRPSTNEOGRTTLYA 407
Db	: :
	342 NTAATDLIPGWHADPRTHSASVSLICKYRLFRGREDLIAGINGYKYAS-NKYGE----- 395
Qy	408 RGGLALNEFSI PQVDLIANARKGVRGYSHTVA-----TENLDDEFI-----YCKST 454
Db	: :
	396 -----RSI-----IPNAIYEFSTKGTAYPOPASFAQTIPOGTRRQITGGYLATR 441
Qy	455 FHPADGLSLTCGGRLGHYKTESGEKGT--LHKASKTKFTGYAGAVYDLADNNNSLYLSLQ 512
Db	: :
	442 FRAADNLSSLILGGRYSRYTGSTQMGTYVVSANRFTPYGTGIVFDLTGNSLYGSYSS 501
Qy	513 LYPTQTNLDADGKLLKPRQNQPEVGYKGSYMDRLNRAVSVFYRMKKDKN---AAPLNPNN 570
Db	: :
	502 LFVPQSKDEHGSLYPKVTCNNLEAGIKGEWLSEGRNLNAGAAVYRARKNLNLAATAAGDPSC 561
Qy	571 KKTRYAALGRVMMEGVETEISGAVTTPKWQIHAGYSYLHSQIKTASNRRDDGIPLLKPKS 630
Db	: :
	562 -NTYYRANQAKTHGWEIEVGGRIITPEWQIQAGYSOKTNRDQGSRINPDVSV-----PERS 616
Qy	631 ANLWTTYQVTPEL-----TIGGGV-----NAMSGITSSAGMHAGGATFED 670

DT 21-MAR-2000 (first entry)
 XX Neisseria meningitidis ORF 742 protein sequence SEQ ID NO:2542.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 XX 11-NOV-1999.
 PD
 XX 30-APR-1999; 99WO-US09346.
 PF
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 XX WPI; 2000-062150/05.
 DR N-PSDB; AA254296.
 DR
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PT
 XX Claim 2; Page 1209; 1453pp; English.
 PS
 XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254673 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 783 AA;
 Query Match 10.5%; Score 397.5; DB 21; Length 783;
 Best Local Similarity 21.3%; Pred. No. 8.6e-27;
 Matches 168; Conservative 92; Mismatches 220; Indels 307; Gaps 27;
 QY 237 YAAADWDINDPTVILGAGVLYQQ-----RHLPYNGLPADANNKLPSPLPQ 280
 DB 3 YGIAEADAGDSVLTGGNYKRSREVPDFSGIILSCENQKTAFFSTP--ACNRPLQLPR 60
 QY 281 HVFVGADWNKFKWNSHDVPADLKHVFGNG-----GYGKVGMRV----- 318
 DB 61 NTYLGEDWSRLSADKYNLPSFGFKHVPDNGWQLNAEVSYTKNESDAKVQOFFLKNPHEAHL 120
 QY 319 SDRDA-----DSNYAFA--G 331
 DB 121 SDEDAVGFLTEKNEVTPPEPKDKALEKLKAYRDETAKEYRERKDDFVKNRPDNTAFQYR 180
 QY 332 SKLGNKTPAGRGCCNTAD---DKACAVGLG-----TEIKQKALA----- 367

Db 191 SRRAERKAGFDECMASAPFALDFICQGSWGDGVDADKSEFVDKALAKEGIFENNAORFP 240
 QY 368 ---FDASYR-----PFR-----LCNTANEFVIGADY--N 392
 Db 241 NSLYDSSFNRRKATARRRYSYMLPRLTKDDRWGIGKLDLTGYTGLPGREHDFVGYAVGDE 300
 QY 393 RFRST-----NEQGR 402
 Db 301 KIRSEYLEIYERRHRVRPNTGATHGVVAGSCQGPDPGDLSPVLRGHKEPDQWQYDEKGN 360
 QY 403 TTYARGGLALNEFRSIPQ-----VDLIANARKGVR----- 433
 Db 361 RTVAEBCRNAKIKTEPKLDAEGKQVYYVDEYSGSRTPVYVDVYELDEKGNKIQETNPD 420
 QY 434 -----GYSHT-----VATENLDEFGIYK-----STFHPADGLSLIGG 466
 Db 421 GTPAFTGSGTVPVWKTVKVADDDHVPALYNYAKYLNNTKTHSLTAGTRFNTVGRHLLOG 480
 QY 467 GRLGHYKIESGEGKTL-----HKASKT---KFTGYAGAVYDLN 501
 Db 481 --LHYTRYETSQTKDMPVRYGQPASDFQTASSIKADQDHYTAKMQGHKLTGYAGITVYDLT 538
 QY 502 DNSSLYLSLSQLYTPQTNLDADGK--LLKPRQNGQFVGYKGSYMDRLNARVSPYRMKDK 560
 Db 539 PQQSIYGSYTKIFKQODNDVDSAKTVLPPLVGTNYEYVWKGAFLOGLRLNASFALFYLEQK 598
 QY 561 N-----AAAPLNPNKKTRYAALGKRVMEGVETEISGAVTPKQIHAQYSYLSHQI 611
 Db 599 NRTVVDVGYVPGAGKQSGPQTVAKPIGKVSRGAEPFELSGELNEDWKVPAGYTYNKSRY 658
 QY 612 KTAS-----NSRDD--GIFLLMPKHSANLWTTYYOV--TPELTIGGVNAMSGITSSA 659
 Db 659 KNAAEVNAERLAKNTGADPNFSNFTPVHIFRFGTSPHPIPTGLTVGGVSAQSGTSSLY 718
 QY 660 GMHAGGYATPDMAAYRFTPKLQIADNINFRHY-----YARVGGANTFNIPOSERTWT 715
 Db 719 NIROGGYGLIDGFVRYELGKHAKLSLIGTLNLRGTYFENNYNRTRGANNFY--GEPRTVS 776
 QY 716 ANLYRSF 722
 Db 777 MKLDWQF 783
 RESULT 13
 ID AAB68923
 XX AAB68923 standard; Protein; 702 AA.
 AC AAB68923;
 XX
 DT 18-APR-2001 (first entry)
 XX
 DE Neisseria meningitidis protein #22.
 XX
 KW Meningococcus; meningitis; bacteraemia; vaccine; dba; fhaB; fhaA;
 KW rni5; rth; tolC.
 XX
 OS Neisseria meningitidis.
 XX
 PN EP1069133-A1.
 XX
 PD 17-JAN-2001.
 XX
 PF 13-JUL-1999; 99EP-0401764.
 XX
 PR 13-JUL-1999; 99EP-0401764.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Nassif X, Tinsley C;
 XX
 DR WPI; 2001-082916/10.

DR N-PSDB; AAF56463.
 XX Immunogenic polypeptides derived from *Neisseria meningitidis* and the
 PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against *Neisseria* infections e.g. bacteraemia and meningitis -
 XX
 XX Claim 3, Fig 22B, 240pp, English.
 XX
 XX The present invention provides the protein and coding sequences of
 CC several genes from *Neisseria meningitidis*. These include the *dsbA*, *fnaB*,
 CC *fhuA*, *rnh15*, *rnh17*, *rnh18*, *rnh19*, *rnh20*, *rnh21* and *tolC* genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.
 XX
 XX Sequence 702 AA;

Query Match 9.9%; Score 375.5; DB 22; Length 702;
 Best Local Similarity 22.2%; Pred. No. 7.3e-25;
 Matches 171; Conservative 136; Mismatches 336; Indels 127; Gaps 33;

QY 8 FRINMTAAATVLAALSSVFAAQTAD-----LEFVHIKQGSRYAIVTEKNGDSSFA 59
 DB 5 FHIALPLTIIA--SEPVAAADTDGNGEHYATLPVSVVGSDTSVLKGYIN--YDEAA 60
 QY 60 VTVGTRIPASLREIPQSVSIITNQVDRNVDTPDQARKTPLRVLSNDGSSVYARG 119
 DB 61 VTRNGQL---IKETPQITDITLNIQKKNVGTNDLSILEGNAGIDAAYDMRG--ESIFLRG 116
 QY 120 YETSEYINI--DGLPAQMSINGTLPNLFAPDRYEVNRGPGSLFDSGEMGCIYVLRKP 177
 DB 117 PQADASDIYRDGVRESGQ--VRSTANI--ERVEILKGPSVLYGRITNGGIVIMVSKYA 172
 QY 178 TKAFQHAAGFGTHQYKAEADVSGSLNSDGSVR-----VMAQTVGASPR 225
 DB 173 NFKQSNIGAVYSGMANRSLIMDINEVLNKNVAIRLTGEVGRANSPRGSDSKVMVSPS 232
 QY 226 PAEKNNRHETFFVAAADWDINDPTVLGAGYLQQORHLAPYNGLPADANMKLPSLPHVFG 285
 DB 233 ITVKLND-----GLKWT-----GQYTYDVERPDRSPKSYVDK--GLPYRNGF-- 276
 QY 286 ADMNKKRNMNSHVF--ADLKHYFGNGGYGKVMRYSDRDADSNYAFAGSKLGKMTPRGPG 344
 DB 277 AHNDVVKVKLQVWRSDILEYAFNDKWRACWQLAHRTAOPDFHFGYASENGNLIKRYAW 336
 QY 345 CNTADKACAVGTEIKQKALAFDASYSRPFILGNTANEFVIGADNRRSRINEGRTT 404
 DB 337 QQT-----DNKTLSSNLTLDGTYTIGREPENHLTVGMDYSR-----EHRNPT 377
 QY 405 LYARGGIALNEFRSIPQVDLIANARKGVRGYSHTVATEN---LDEFGYKSTFHPADGL 461
 DB 378 LGRVRSF-----TVF--INPYDRAWMPASGRLOPILQNHKADSYGIFQNIFFSATPDL 430
 QY 462 SLIGGRGLHYKTESGEGKTLHKASKTKETGYA-----GAYVDLNDNSIYLSLSQLYTP 516
 DB 431 KFLVGGRYKTYTNSENKLT---GNSRQYSGHSFSPNIGVWMININPHLTLYASYNKGFAP 487
 QY 517 -----GTNLDAQDKL---KPRGNGQFVGVKGSYMDDLNANVSFPMKDKRAAALNP 568
 DB 488 YGGRGLSIDTSSAAVFNAAPEYTRQETGVKSSWDDRLSTLLSAQIERFIRYRDP 547
 QY 569 NNKTRVYALGKRVMEGVETETISGAVTPKQIHAQSYLSLSQIKTASNSRHD---GIFLL 625
 DB 548 KNNPIYIYASGKRRSGVELSAIGQIIPK--KLYLRGSLGVMQAKYVEDKENPRVGIHLN 606
 QY 626 MPRH--SANTMTTYQVTPELTIGGVANMSGITSSAGNHAG-----GYATFDAM 672
 DB 607 NTSNVTGNLFFRY--TPTENLYGEI-----GVGTG--GKRYGNSRNKEVTLLPFAFARDAM 659
 QY 673 AARFTRPKLQIINADNIPRRHYAVRGANTNINPGESEPTWANKRYSF 722
 DB 660 LGNNH--KNVAVTTPAANLNLQKYM-----RSDSPGNGPRGYTARAVYRFR 702

RESULT 14

AAB68921

ID AAB68921 standard; Protein; 702 AA.

XX

XX AAB68921;

XX

XX 18-APR-2001 (first entry)

XX

XX *Neisseria meningitidis* protein #20.

XX

XX Meningococcus; meningitis; bacteraemia; vaccine; *dsbA*; *fnaB*; *fhuA*;
 KW *rnh15*; *rnh*; *tolC*.
 XX

XX

XX *Neisseria meningitidis*.
 OS

XX

XX EPI069133-A1.

XX

XX 17-JAN-2001.

XX

XX 13-JUL-1999; 99EP-0401764.

XX

XX 13-JUL-1999; 99EP-0401764.

XX

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (PLAC) MAX PLANCK GBS FOERDERUNG WISSENSCHAFTEN.
 XX

XX

XX Nasself X, Tinsley C;
 PIXX WPI; 2001-082916/10.
 DRXX N-PSDB; AAF56461.
 XX

XX

XX Immunogenic polypeptides derived from *Neisseria meningitidis* and the
 PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against *Neisseria* infections e.g. bacteraemia and meningitis -
 XX

XX

XX

XX

XX Claim 3, Fig 20B, 240pp, English.

XX

XX The present invention provides the protein and coding sequences of
 CC several genes from *Neisseria meningitidis*. These include the *dsbA*, *fnaB*,
 CC *fhuA*, *rnh15*, *rnh17*, *rnh18*, *rnh19*, *rnh20*, *rnh21* and *tolC* genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.
 XX

XX

XX

XX

XX

XX

XX Sequence 702 AA;

XX

Query Match 9.9%; Score 374.5; DB 22; Length 702;
 Best Local Similarity 22.5%; Pred. No. 9e-25;
 Matches 175; Conservative 134; Mismatches 326; Indels 143; Gaps 35;

QY 8 FRINMTAAATVLAALSSVFAAQTAD-----LEFVHIKQGSRYAIVTEKNGDSSFA 59
 DB 5 FHIALPLTIIA--SEPVAAADTDGNGEHYATLPVSVVGSDTSVLKGYIN--YDEAA 60
 QY 60 VTVGTRIPASLREIPQSVSIITNQVDRNVDTPD--OLAKTPLGLRVLSNDGSSVYA 117
 DB 61 VTRNGQL---IKETPQITLTL--NIKKNVGTNDLSILEGNAGIDAAYDMRG--ESIFL 113
 QY 118 RGYEYSEYINI--DGLPAQMSINGTLPNLFAPDRYEVNRGPGSLFDSGEMGCIYVLRKP 175
 DB 114 RGFQADASDIYRDGVRESGQ--VRSTANI--ERVEILKGPSVLYGRITNGGIVIMVSK 169
 QY 176 RPTKAFQHAAGFGTHQYKAEADVSGSLNSDGSVR-----GR-----VMAQTVGAS 223
 DB 170 YANFKQSRIGAVYSGMANRSLIMDINEVLNKNVAIRLTGEVGRANSPRGSDSKVMVSK 229
 QY 224 PRPAEGNNRHETFFVAAADWDINDPTVLGAGYLQQORHLAPYNGLPADANMKLPSLPHVFG 283
 DB 230 PSLITVKLND-----GLKWT-----GQYTYDVERPDRSPKSYVDK--GLPYRNG 274
 QY 284 VGADWNRKRNNSHVF--ADLKHYFGNGGYGKVMRYSDRDADSNYAFAGSKLGKMTPRG 342

Db 275 F-AHRNDFVKDKLQVRSDEYAFNDKWAQOHLAHTAAQDFHYPAGSENGNLIKRYN 333
 Qy 343 PGCNTADDKACAVGLTEIKQKALAFDASYSRPFRLGNTANFVIGADYNRFRSTNEQGR 402
 Db 334 ANQQT-----DNKTLSSNFTLNGDITIGREPHNLTVGMDYSR-----EHRN 374
 Qy 403 TTYLARGGLALNEFRSIPQVDLIANARKGVGYSHTVATEN---LDEFGYKSKTFHPAD 459
 Db 375 PTLGVRGSP-----TVP-INPYDRASWPASGRLOPILTONRHKADSYGIFVQNIPTATP 427
 Qy 460 GLSLGGGRGLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNDNNLSYLSLSOLY 514
 Db 428 DLKFLVGGRYDKYTFENSEKLT---GNSQYSGHGSFSPNIGAVWNPVHTLIYASYNKGF 484
 Qy 515 TP-----QTNLDAGKLLKPRQGNQFEVGYKSGYMDRLNARVSYFYRMKDK 560
 Db 485 APYGGRGYLSIDTLSSAVFNAD---PEYTRQYETGVKSSWLDRLSTLTSVAQIERP 539
 Qy 561 NAAAPLNPNNKTRYAALGKRVWEGVETISGAVTPKQIHAGYSYLHSQIKTASNSRDD 620
 Db 540 NIRYRDPKNNPIYAVSGKHSRGVELSAIGCIIPK-KLYLRGSLGVNQAKVVEDKENP 598
 Qy 621 ---GIFLLMPKH-SANLWTTYQVTPELTIGGGVYAMSGITSSAGMHAG----- 664
 Db 599 DRVGHLNNTSVNIGLFFRY--TPTENLYGEI-----GVGTG-GRYGVNSRNKEVTTLP 651
 Qy 665 GYATDAMAARYFTPKLQKIQINADNIFNRHYARVGGANTFNIPGSERTWTANLRYSF 722
 Db 652 GFARVDAMLGWNH-KNVNVTFAANLFNQYK-----RSDSMPCNPGRYTARVNYRF 702

 RESULT 15
 ABB52462
 ID ABB52462 standard; Protein; 663 AA.
 XX
 AC ABB52462;
 DT 11-FEB-2002 (first entry)
 DE
 DE Escherichia coli polypeptide SEQ ID NO 263.
 XX
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicemia;
 KW pyelonephritis; antibiotic resistance.
 XX
 OS Escherichia coli.
 XX
 PN WO2001166572-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 12-MAR-2001; 2001WO-EP03445.
 XX
 PR 10-MAR-2000; 2000FR-0003145.
 PR 02-FEB-2001; 2001FR-0001449.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 XX
 DR WPI; 2001-550253/61.
 XX
 PT A library of DNA fragments of Escherichia coli strains for the
 PT phylogenetic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A- -
 XX
 PS Example 6; Fig 6; 646pp; English.
 XX
 CC The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
 CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
 CC B2/D+A-. The polynucleotides have potential antiinflammatory.

CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.
 XX

SQ Sequence 663 AA;
 Query Match 9.8%; Score 371; DB 22; Length 663;
 Best Local Similarity 22.1%; Pred. No. 1.7e-24;
 Matches 171; Conservative 121; Mismatches 285; Indels 198; Gaps 32;

Qy 20 ALSSSVFAA-----QTADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGKIPASL 70
 Db 15 ALFELLFAAPMIHATDSVTYTKDGETIIVTADAN---TATEATDGYQLSTSTATLTDMPM 71
 Qy 71 REIPQSVSIITNQOVKDRNVDTPDLARKTPGLRVLSNDDG--RSSVYARGYSEYNID 128
 Db 72 LDIPQVNTVSDQVLEQNQATTLDEALYVNSNV-VQNTILGGTQDAFVRGTF---GANRD 127
 Qy 129 GLPQMQSINGTLNLPF--AFDRVEMRGSGLEFDSSGEMGGIVNLRKPTKAFQGHAA 186
 Db 128 G-SIMTNGRLTVLPSRNAATERVELKGPASTLYGLIDPGLINVVYTKPEKTFHGSVS 186
 Qy 187 A-----GFGTHKQYKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEK---NNRHETFY 237
 Db 187 ATSSSFGGT-----GQDITGTEGTQLAYRLTGEVQDEDIYWRNFGRKSTFI 235
 Qy 238 A-AADWDINPDVILGAGLYQQRHLAPYNGLPADANNKLPSPQHFVFGADWNKFKMNSH 296
 Db 236 APSLTFQGDNATVTMAEY-----HL-----NSQWT----- 260
 Qy 297 DVFADLKHFGNGGYGKVMRYSDRSDADSNVAFAGSKLGMKTPAGRPGCNTADDKACAVG 356
 Db 261 -----ARFDYSYSDQKY-----SDNQA----- 277
 Qy 357 LGTEIKOKALAFDASYSRPPF-----LGNTANEFVIGADYNRFRSTNEQGRITLYA--- 407
 Db 278 -----RVYAYDATTGTLTRVDATOGSTQRMHATRADL-----QGNVDIAGFYNE 322
 Qy 408 -RGGLALNEFRSIPQVDLIA-----NARKGVRGYSHTVATENLDE-----FCIY 450
 Db 323 LLGGVSY-EYVDLLRTDMIRCKKAKDFNIYNPVYNTSKCTTVSASDSQTIKQENYSAY 381
 Qy 451 GKSTFHPADGLSLGGRLGHYKIESGEGKTLHKASKTK---FTGYAGAVYDLNDNNLSY 507
 Db 382 AQDALYLTDMNIAVAGIRYQYTYQYAGKGRPFNVNTDSRDEQMTPKLGLVYKLTTPSVSLF 441
 Qy 508 LLSQLYTPQNLDAADGKLLKPRQGNQFEVGYKSGYMDRLNARVSYFYRMKDKNAAAPLN 567
 Db 442 ANYQTMPQSSIASYIGDLPPSSNAYEYQAKFELFDG-ITADIALFDIHKRNLVLTES 500
 Qy 568 PNNKTRYAALGKRVMEGVETEISGAVTPKQIHAGYSYLHSQI-----KTASNSRD 619
 Db 501 IGDETIAKTA-GRVRSRGVEVDLAGALTENINIIASGYTDAKVLDEPDYAGKPLPN--- 556
 Qy 620 DGIFLLMPKHSANLWTTYQV-----TPELTIGGGVYAMSGITSSAGMHAGGYATFDAM 672
 Db 557 -----VPRHTGSLFTYDIHNNMFGNNTLTFGGGGHGVSRSSATNGADYLYLPGYFVADAF 610
 Qy 673 AAYRFTPKLK-----LQINADNIFNRHY-ARVGGANTFNIPGSERTWTANLRYSP 722
 Db 611 AAYKM--KLQYPTVLQNLVKNLFDKTYTTSIATNNLGNQIGDPRVQFTVKMEF 663

RESULT 16
 AAB68926
 ID AAB68926 standard; Protein; 703 AA.
 XX

```

AC AAB68926;
XX
XX 18-APR-2001 (first entry)
XX
XX Neisseria meningitidis protein #25.
DE
XX
XX Meningococcus; meningitis; bacteraemia; vaccine; dba, fnaB, fnaH;
KM rml5, rth; toIC.
XX
OS Neisseria meningitidis.
XX
XX EPI069133-A1.
XX
XX 17-JAN-2001.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX (INM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Nassif X, Tinsley C;
XX
XX WPI; 2001-082916/10.
XX
XX N-PSDB; AAF56466.
XX
XX Immunogenic polypeptides derived from Neisseria meningitidis and the
XX nucleic acids that encode them, useful for diagnosing and vaccinating
XX against Neisseria infections e.g. bacteraemia and meningitis -
XX
XX Claim 3; Fig 25B; 240p; English.
XX
XX The present invention provides the protein and coding sequences of
XX several genes from Neisseria meningitidis. These include the dba, fnaB,
XX fnaH, rml5, rth17, rth18, rth19, rth20, rth21 and toIC genes. These can
XX be used in the diagnosis and treatment of infection by the bacterium,
XX which can lead to meningitis and bacteraemia, and in vaccines to prevent
XX such infection.
XX
XX Sequence 703 AA;
SQ
Query Match 9.8%; Score 371; DB 22; Length 703;
Best Local Similarity 22.4%; Pred. No. 1.9e-24;
Matches 174; Conservative 131; Mismatches 332; Indels 140; Gaps 33;
QY 8 PRINMATAIVLAASSVPAQAQAD-----LETVHKGQSNVAIYERKGDYSSRA 59
DB 5 PHLALPLTLIA--SEPVAAADTDQNGEHYATLPTVSVVGSDTSVLKGYN--YDEAA 60
QY 60 VTVGTRIPASLREIPQSVSITITQGVKRVNDFDQIARKEPGLRVLSNDDGRSSVYARG 119
DB 61 VTRNGD---IKETPOTITLNIQKKNKGTNDLSLSEGNAGIDAIYDRG-ESIFLRG 116
QY 120 YEYSEYNI--DGLPAQMSINGTLFNLFAFDYEVNRGSPGLDSSGEMGIVNLVRKEP 177
DB 117 FQADADIVRDGVRSEGC-VRRSTANI---ERYEILKGPSSVLVGRITGGGVINMYSKA 172
QY 178 TYAFQGHAAAGRTHKQYRAEADVSGSLNSDSVR-----VMAQTVGASPR 225
DB 173 NFKQSHNIAVVGSMNRSLMDINELINKVAIRLTGEVGRNSFRSGIDSKNVVSPS 232
QY 226 PAKKNRHETFFVAAADMDINPDTVLGAGYLQOGRHLAPYNGLPADANNKLPSPQVFPVG 285
DB 233 IYVKLDN-----GLKKT-----GQITYDVERTPDRSPFKSYDR-GLPYMGF- 276
QY 286 ADMNKEKRNASHVF-ADLRHYFGNGGYGVKMYSDRDADSNYAFAQSKLGKMTKTPGRPG 344
DB 277 AHNNDVFKKLQVWRSDLEAYAFNDKWRAGWQLAHRTAQQDFHFVYAGSEGNLIKRVYAW 336
QY 345 CNTADAKACAVGLGTETIKQKALAFDASYSRPFILGNTAFVYGAAYNRNRSTNBOGRTT 404
DB 337 QQT-----DNKTLSSNFTLNGDYITIGRFENHLTVGMDYSR-----EHRNPT 377

```

```

QY 405 L-YARGGLALNEFRSIPVDLIANARKGVRSHTVATEN---LDEFGIYGSSTPHADG 460
DB 378 LGYSRAFTA-----SIDPYDRASWPASGRLOPILTONRHKADSYGIFQNIFFATPD 429
QY 461 LSLIGGRGLGHKIBESBEKTLHKASKTFGTGA-----GAYVDLNDNNSLYLSLSQLYT 515
DB 430 LKFVLGGRYDKTYFNSENKLT---GSSRQYSGHSFSPNIGAWNINPVTLLVASYAKGPA 486
QY 516 P-----QTNLDADGKLLKPRQNOFEVGYGSYMDDRLNARVSFPRMKDN 561
DB 487 FYGGRGGYLSTSSSAVNAVAD-----PEYTRYETGVKSSWDDRLSTLLSAIQIERN 541
QY 562 AAAPLNPNKKTRTYALGRVMEGVETEISGAVTPRQTHAGSYLSHQITKASNSRDD- 620
DB 542 IRYRPDQNDPYTMAVGGHRSRGVELSAIGQIIPR-KLYLNGSLGVMAQKVEDKENPD 600
QY 621 --GIFLMPKH-SANLWTTYQVTPELTIGGVNANSGITSSAGMAG-----G 665
DB 601 RVGIHLNNTSNVTGNLFPRY--TPTENLYGEI---GVLTG-GRKRYGNSRKEVTTLPG 653
QY 666 YATFDMAAYRFTPKLQIINDNIFNRHYARVVGANTFNI PGSERTWTANLRYSF 722
DB 654 FARVDMLGWNH-KNVNITFPAAANLNLQKW-----RSDAMPGARPTTARVNTSF 703
RESULT 17
AA75568
ID AA75568 standard; Protein; 703 AA.
XX
XX AA75568;
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 761 protein sequence SEQ ID NO:2610.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
XX
XX WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX
XX 31-JUL-1998; 98US-0094869.
XX
XX 02-SEP-1998; 98US-0098994.
XX
XX 02-SEP-1998; 98US-0098062.
XX
XX 09-OCT-1998; 98US-0103749.
XX
XX 09-OCT-1998; 98US-0103794.
XX
XX 09-OCT-1998; 98US-0103796.
XX
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHTR ) CHIRON CORP.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Veneri JC;
XX
XX WPI; 2000-062150/05.
XX
XX N-PSDB; AA254330.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX
XX Claim 2; Page 1238; 1453p; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
XX
XX

```


QY 345 CNTADKACAVGLGTEIKQKALAFDASYSRPRFLGNTANEFYGAATYNNFRSTNEGRTT 404
 Db 337 QQT-----DKTLSSNLTLLNGDYITIGRFENHLTVGMDSYR-----EHRNPT 377
 QY 405 LYARGGLALNEFRSIPQVLLINARKGVAGYSHVATEN---LDERGIYKSTFHPADGL 461
 Db 378 L----GFSSAFSASINPYD---RASWPASGRLOPILTONRHRADSGIGFQNIFFSATPDL 430
 QY 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNNSLYLSLSOLYTP 516
 Db 431 KFYLGGRYKTYTNSENKLT---GSSROYSGHSPFNIGAVNNINVHLTYASYNNGFAP 487
 QY 517 -----QTNLDADGKLKPRGQNOPEVGYKSYWDDRLNARVSFRMKDKKA 562
 Db 488 YGGRGGYLSIDTLSSAVFNAD-----PEYTRYETGVKSSWLDRLSTLLSAVQIERFNI 542
 QY 563 AAPLNPNKKTRYAALGKRVMEGVETEISGAVTPKQIHAGSYLHSQTKTANSNDD-- 620
 Db 543 RYRPDPKNNPYIYAASGKRSRGVELSAIGQIIPK-KLYLRGSLGVMQAKVVEDKENPDR 601
 QY 621 -GIFLLMPKH-SANLWTTYQVTPELTIGGVNAMSGITSSAGMHAG-----GY 666
 Db 602 VGHLNNTSNVGNLFERR--TPTENLYGEI---GVGTG-GKRYGYSRNNKEVTTLPDF 654
 QY 667 ATFDAMAAYRTPPKLKLQINADNIFNRHYAARVGANTNIPGSEBRTWTANLRYSF 722
 Db 655 ARVDAMLGNWH-KNVAVTFPAAANLNFQKTYW-----RDSMPGNPRGYARVAVYRF 703

RESULT 19

AA75567
 ID AA75567 standard; Protein; 703 AA.

AC AA75567;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 761 protein sequence SEQ ID NO:2608.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy.

OS Neisseria meningitidis.

PN MO9957280-A2.

PD 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 09-OCT-1998; 98US-0103796.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;

XX Pieterzen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX Tettelein H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; AA54329.

XX Novel Neisserial polypeptides predicted to be useful antigens for

XX vaccines and diagnostics

PS Claim 2; Page 1237; 1453pp; English.
 XX AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 703 AA.

Query Match 9.7%; Score 368; DB 21; Length 703;

Best Local Similarity 22.4%; Pred. No. 3.5e-24;

Matches 174; Conservative 130; Mismatches 334; Indels 138; Gaps 33;

QY 8 FRINMTAAVTLAASSVSAAQTAD-----LEVNHKQGRSTVAITEKQDYSRA 59
 Db 5 FHIALPLTLIIA--SPVPAADTQDGEHYATLPTVSVVGQSDISLKYIN--YDEAA 60
 QY 60 VYVGTGIPASLEIPQSVSIITNQGVKDRNVDTPDQAKRTGLRYLSNDDGRSSVYARG 119
 Db 61 VTRNGQL---IKETPOTITLINIQKKNYGTNDLSILEGNAGIDAAVYMRG-ESIFLKG 116
 QY 120 YEYSEYNI--DGLPAQMGISNGTLPNLFAPDREYVWRGSGSFDSSGEGIVNLVRKP 177
 Db 117 FQADAGDIYRDGVRESGQ--VRRSTANI---ERVELIKGSSVLYGRNCGGVYNNVSKYA 172
 QY 178 TKAFQGHAAAGFTKQYARADVSGSLNSDGSVR-----VMAQTVASRR 225
 Db 173 NFKQSRNIGAVYGSWANRSLNMDINEVLKNVAIRLTGEVGRANSFRSGIDSKNVAWSPS 232
 QY 226 PAEKNRHEFTFYAADWDINDPVLGAGYLYOQRHLAPNGLPADANNTLPSPQVFPVG 285
 Db 233 IYVKLDN-----GLKMT-----GQYTYDNVERTTPRSPTKSYVDYF-GLPYRMGF- 276
 QY 286 ADWNRKPKNSHDVF-ADLKHVFGNGGYGVKMYSDRDADSNVAFASGKLGMKTPARGRG 344
 Db 277 AHRNDFVKDQLVWMSDLEAYAFNDKWRQAQWQLAHRPAQDPHFVYAGSGENGLIKENYAW 336
 QY 345 CNTADKACAVGLGTEIKQKALAFDASYSRPRFLGNTANEFYGAATYNNFRSTNEGRTT 404
 Db 337 QQT-----DKTLSSNLTLLNGDYITIGRFENHLTVGMDSYR-----EHRNPT 377
 QY 405 LYARGGLALNEFRSIPQVLLINARKGVAGYSHVATEN---LDERGIYKSTFHPADGL 461
 Db 378 L----GFSSAFSASINPYD---RASWPASGRLOPILTONRHRADSGIGFQNIFFSATPDL 430
 QY 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNNSLYLSLSOLYTP 516
 Db 431 KFYLGGRYKTYTNSENKLT---GSSROYSGHSPFNIGAVNNINVHLTYASYNNGFAP 487
 QY 517 -----QTNLDADGKLKPRGQNOPEVGYKSYWDDRLNARVSFRMKDKKA 562
 Db 488 YGGRGGYLSIDTLSSAVFNAD-----PEYTRYETGVKSSWLDRLSTLLSAVQIERFNI 542
 QY 563 AAPLNPNKKTRYAALGKRVMEGVETEISGAVTPKQIHAGSYLHSQTKTANSNDD-- 620
 Db 543 RYRPDPKNNPYIYAASGKRSRGVELSAIGQIIPK-KLYLRGSLGVMQAKVVEDKENPDR 601
 QY 621 -GIFLLMPKH-SANLWTTYQVTPELTIGGVNAMSGITSSAGMHAG-----GY 666
 Db 602 VGHLNNTSNVGNLFERR--TPTENLYGEI---GVGTG-GKRYGYSRNNKEVTTLPDF 654
 QY 667 ATFDAMAAYRTPPKLKLQINADNIFNRHYAARVGANTNIPGSEBRTWTANLRYSF 722
 Db 655 ARVDAMLGNWH-KNVAVTFPAAANLNFQKTYW-----RDSMPGNPRGYARVAVYRF 703

CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.

XX Sequence 702 AA;

Query Match 9.7%; Score 367.5; DB 22; Length 702;
 Best Local Similarity 22.6%; Pred. No. 3.9e-24;
 Matches 176; Conservative 131; Mismatches 327; Indels 145; Gaps 35;

```

QY 8 FRINMTAATVLAALSSVPAQAOTAD-----LETVHIKQGRSNAIVTEKNGDYSSFA 59
DB 5 FILALPLTIIIA--SFPVAADTODNGEHYATLPTVSVVGSDTSVLKGYIN--YDEAA 60
QY 60 VTGVTKIPASLREIPQSVSIITNQVKDRNVDTFD--QLARKTPGLRVLSDNGRSSVYA 117
DB 61 VTRNGQL---IKETPQTIDTL---NIKNNKYGINDLSILEGNAIDAYDMRG-ESIFL 113
QY 118 RGEYSEYNI--DGLPAQMSINGTLPNLFAPDRVEVMGPGSLFSSGEMGIVLVLRK 175
DB 114 RGFQADASDIYRDGVRESQ--VRRSTANI---ERVELIKGPPSVLYGRITNGGIVNWSK 169
QY 176 RPTKAFQGAAGFTGHKQYKAEADVSGSLNSDGSVR-----VMAQTVGAS 223
DB 170 YANFKOSRITGAVYGSWNRSLNMDINEVLANKVAIRLTGCEVRANSFRSGIDSKVWVS 229
QY 224 PPAEKNNRHETFFYAADWDINPTVLGAGYLYQORHLAPYNGLPADANNKLPSLPQVFG 283
DB 230 PSITVLDN-----GLKMT-----GQYTYDNVERTPDRSPKTSYVDRF-GLPYRMG 274
QY 284 VQADNNKFKQNSHDVF-ADLKHYFGNGGYGKVMRYSDDRADSNYAFAGSKLGMKTPAGRP 342
DB 275 F-AHRNDFYVYDKLQVWRSDLEYVAFNDKWRARQWOLARTAAODPHFYAGSENNLTKRY 333
QY 343 PGONTADKACAVGLTETIKOKALAPDASYSRPFRLGNTANEVADYVRFSTNEQGR 402
DB 334 AANQOT-----DNKTLSSNFTLNGDYTTGRENNHLLTVGMDSR-----EHRN 374
QY 403 TTL-YARGGIALNEFRSIPQVDLIANARKGVRSHTVATEN--LDEFGIYKSTFHPA 458
DB 375 PTLGYSRAFTA-----SIDPYRASNPASGRLPILTONHKAIDSYGIFQNFIFSAT 426
QY 459 DGLSLIGGRGLHYKTESGEGKTLHKAISKTFGYA-----GAYTLDNNSLYLSLSQL 513
DB 427 PDLKFLVGRKYDTFNSENKLT---GSSRQYSGHSFSPNIGVWNNINPVHTLYASYNKG 483
QY 514 YTP-----QTNLDADGKLTKPRQNOFEGYKGSYMDRLNARSFPRMD 559
DB 484 FAPYGGRGYGLSTINSSSAVFEND---PEYTRYETGVKSSWLDRLSTLTAAYQIER 538
QY 560 KNAAPLNNNNKTRYYAALGKRVMEGETEISGAVTPKMQIHAGSYLSQIKTAANSRD 619
DB 539 FNIIRYAPDQNDPYTAVAGSKHRSRGVELSAIQIIFK-KLYIRGSLGVWQAVYVDKEN 597
QY 620 D--GIFLMPKH-SANLMTTYQVTELTITGGGVNAMSGITSSAGMHAG-----664
DB 598 PRVGVHILNNTSVTGNLFFRY--TPTENLYGSI---GVYGT-GRKGYNSHKEVTTL 650
QY 665 -GAYTDAAMAARFTPKLKLQINADNIENRHYYARVGANTFNIIPSESRWTANLAYSF 722
DB 651 PGFARVADLGMWH-KNVNITTPAANLNLQKYM-----RSDAMPAPRTYATVAVYSF 702

```

RESULT 22
 AAB18720
 ID AAB18720 standard; Protein; 703 AA.

XX AAB18720;

DT 22-JAN-2001 (first entry)

XX A Neisseria meningitidis BASB083 polypeptide.
 DE BASB082; BASB083; BASB091; BASB092; BASB0101; infection; vaccine;

KW gene therapy; upper respiratory tract infection; bacteraemia; meningitis;
 KW invasive bacterial disease.

XX Neisseria meningitidis.

XX MO200055327-A2.

XX 21-SEP-2000.

XX 07-MAR-2000; 2000MO-EP01955.

XX 12-MAR-1999; 99GB-0005815.

XX 21-APR-1999; 99GB-0005094.

XX 23-APR-1999; 99GB-0005003.

XX 28-APR-1999; 99GB-0009787.

XX 07-MAY-1999; 99GB-0010710.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Defrenne C, Delmelle C, Ruelle J;

XX WPI: 2000-602119/57.

XX N-PSDB; AAA75745.

XX Novel polypeptides designated BASB 082, 083, 091, 092, and 101 derived

XX from meningococcus bacterium useful for producing vaccines against

XX infections and in diagnostic assays -

XX Claim 3; Page 102-104; 108pp; English.

The present sequence represents a BASB083 polypeptide. The specification describes BASB082, BASB083, BASB091, BASB092, and BASB0101 polypeptides. The polynucleotides and polypeptides are useful as diagnostic reagents and for diagnosing N. meningitidis infection. The polynucleotides may be used as hybridisation probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding BASB082, BASB083, BASB091, BASB092 or BASB0101 polypeptides and to isolate cDNA and genomic clones of other genes that have a high identity particularly high sequence identity to BASB082, BASB083, BASB091, BASB092 or BASB0101 genes. The CC vaccine compositions are useful for inducing an immunological response in humans. The polynucleotides encoding BASB082, BASB083, BASB091, CC BASB092 or BASB0101 polypeptides are useful in gene therapy to induce an immunological response. The polypeptides are useful for treating CC upper respiratory tract infection, invasive bacterial diseases, such as CC bacteraemia and meningitis.

XX Sequence 703 AA;

Query Match 9.7%; Score 366; DB 21; Length 703;

Best Local Similarity 22.2%; Pred. No. 5.4e-24;
 Matches 172; Conservative 133; Mismatches 333; Indels 138; Gaps 33;

```

QY 8 FRINMTAATVLAALSSVPAQAOTAD-----LETVHIKQGRSNAIVTEKNGDYSSFA 59
DB 5 FILALPLTIIIA--SFPVAADTODNGEHYATLPTVSVVGSDTSVLKGYIN--YDEAA 60
QY 60 VTGVTKIPASLREIPQSVSIITNQVKDRNVDTFDQLARKTGLRVLSDNGRSSVYAR 119
DB 61 VTRNGQL---IKETPQTIDTLNIQKNNKYGINDLSILEGNAIDAYDMRG-ESIFLNG 116
QY 120 YEYSEYNI--DGLPAQMSINGTLPNLFAPDRVEVMRGPGLFSSGEMGIVLVLRKP 177
DB 117 FQADASDIYRDGVRESQ--VRRSTANI---ERVELIKGPPSVLYGRITNGGIVNWSKA 172
QY 178 TKAFOGHAAGFTGHKQYKAEADVSGSLNSDGSVR-----VMAQTVGASPR 225
DB 173 NFKOSRITGAVYGSWNRSLNMDINEVLANKVAIRLTGCEVRANSFRSGIDSKVWVS 232
QY 226 PPAEKNNRHETFFYAADWDINPTVLGAGYLYQORHLAPYNGLPADANNKLPSLPQVFG 285
DB 233 ITVYKLDN-----GLKMT-----GQYTYDNVERTPDRSPKTSYVDRF-GLPYRMG- 276
QY 286 ADWNNKFKQNSHDVF-ADLKHYFGNGGYGKVMRYSDDRADSNYAFAGSKLGMKTPAGRP 344

```

```

Db 277 AHRNDFVKDLQWRSDELEYAENDKWAQOLAHRTAAQDFHIFYAGSENGNLIKRNAY 336
QY 345 CNTADDKACAVGLGTEIKQALAFDASYSRPRLGNANTANEFVIGADYNFRSTNQGR 404
Db 337 QQT-----DNKTLSSNFTLNGDYITGRFENHLTVGMDYSR-----EHRNPT 377
QY 405 LVARGGLALNEFRSIPQVDLIANARKGVGYSHTVATEN---LDFEIGYKSTFHPADGL 461
Db 378 L-----GFRNFASIDPYD---RASRPASGRQLQILAQDRHKADSYGIFVQWIFSATPDL 430
QY 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----CAVDLDNNSLSLSLSQLYTP 516
Db 431 KPVLGGRYDKYTFNSENKLT---GSSRQYSGHSPNIGAVNNINPVHTLYASYNAFAP 487
QY 517 -----QTNLDADGKLLKPRQGNQFQEVGYKGSYMDRLNARVSFYRMKDKN 562
Db 488 YGGRGGYLSINTSSSAVFNAD-----PEYTRQYETGVKSSWLDRLSTLSAYQIERNI 542
QY 563 AAPLPNNKTRYAALGKRVMEGVETEISGAVTPKWQIHAGYSYLHSQIKTASNRDD-- 620
Db 543 RYRPDEQNDPYTWAVGGKHSRSGVELSAIQIIPK-KLYLRGSLGVQAKVVEDKKNPDR 601
QY 621 -CIFLLMPKH-SANLWTTTYQVTPBELTIGGVNAMSGITSSAGMHAG-----GY 666
Db 602 VGIHLNNTSNVTGNLFRRY--TPTENLYGEI---GVGTG-KGRYGYNSRNKEVTTLPGF 654
QY 667 ATFDAMAAVRFPTKCLKLOINADNIFNRHYVYRVGGANTFNIPGSERTWTANLYSF 722
Db 655 ARVDAMLGNWH-KNVNVTFAANLFNQKW-----RSDSMFGNPRGYTARVNYRF 703

```

RESULT 23

```

AAB68922
ID AAB68922 standard; Protein; 702 AA.
XX
AC AAB68922;
XX
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis protein #21.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dba; fhaB; fhuA;
KW rni5; rth; toIC.
XX
OS Neisseria meningitidis.
XX
PN EP1069133-A1.
XX
PD 17-JAN-2001.
XX
PF 13-JUL-1999; 99EP-0401764.
XX
PR 13-JUL-1999; 99EP-0401764.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Nassif X, Tinsley C;
XX
WPI; 2001-082916/10.
DR N-PSDB; AAF56462.
XX

```

Immunogenic polypeptides derived from *Neisseria meningitidis* and the nucleic acids that encode them, useful for diagnosing and vaccinating against *Neisseria* infections e.g. bacteraemia and meningitis -

Claim 3; Fig 21B; 240pp; English.

The present invention provides the protein and coding sequences of several genes from *Neisseria meningitidis*. These include the *dba*, *fhaB*, *fhuA*, *rni5*, *rth17*, *rth19*, *rth20*, *rth21* and *toIC* genes. These can be used in the diagnosis and treatment of infection by the bacterium,

CC which can lead to meningitis and bacteraemia, and in vaccines to prevent such infection.

SQ Sequence 702 AA;

Query Match 9.7%; Score 364.5; DB 22; Length 702;
 Best Local Similarity 22.8%; Pred. No. 7.3e-24;
 Matches 177; Conservative 129; Mismatches 330; Indels 141; Gaps 35;

QY 8 FRINMTATVLAALSSSVFAAQAD-----LETVHIKQQRSYNAIVTEKNGDYSSFA 59

Db 5 FHLALLPLIIA--SFPVAAADTDNGEHYATLPTVSVVGQSDTSVLKGIN--YDEAA 60

QY 60 VTVGTKIPASUREIPQSVSIITNQVKDRNVDTFDQLARKTPGLRVLNSDDGRSSVYARG 119

Db 61 VTRNGQL---IKETPQTIDTLNIQKNKNGYTNLDSSILEGNAGIDAAAYDMRG-ESIFLRG 116

QY 120 YEYSEYNI--DGLPAQMQSINGTLPNLPAPDRVEMRGPSGLFDSSGEMGGIVNLVRKP 177

Db 117 FOADASDIYRGVRESGQ-VRRSTANI---ERVELKGPSSVLVYGRNTGGGVINMVKYA 172

QY 178 TKAFQGHAAAGFGTHKOYKAEDVSGSLNSDGSVR-----GR-----VMAQTVGASPR 225

Db 173 NFKOSRNIGAVYGWNR-SLNMDINEVLNKNVAIRLTCEVGRANSFRSGIDSKNMVSPS 231

QY 226 PAEKNRHETFYAAADWDINDPTVLGAGYLYQORHLAPYNGLPADANNKLPSPQHFVVG 285

Db 232 ITVKLDN-----GLKWT-----GQTYDNDVERTPDRSPTKSVYDRF-GLPYRMGF- 275

QY 286 ADMNKFKNSHDVF-ADLKHVFGNGYKVGMRYSDDADSNYAFAGSKLGMKTPAGRPG 344

Db 276 AHRNDFVKDKLQWRSDELEYAFNDKWAQOLAHRTAAQDFHIFYAGSENGNLIKRNAY 335

QY 345 CNTADDKACAVGLGTEIKQALAFDASYSRPRLGNANTANEFVIGADYNFRSTNQGR 404

Db 336 QQT-----DNKTLSSNFTLNGDYITGRFENHLTVGMDYSR-----EHRNPT 376

QY 405 L-YARGGLALNEFRSIPQVDLIANARKGVGYSHTVATEN---LDFEIGYKSTFHPADG 460

Db 377 LGYNRAFSA-----SINPYD---RASWPASGRQLQIIPK-KLYLRGSLGVQAKVVEDKKNPDR 428

QY 461 LSLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLDNNSLSLSQLYT 515

Db 429 LKFLVGGRYDKYTFNSENKLT---GSSRQYSGHSPNIGAVNNINPVHTLYASYNKGA 485

QY 516 P-----QTNLDADGKLLKPRQGNQFQEVGYKGSYMDRLNARVSFYRMKDKN 561

Db 486 PYGGRGGYLSIDTSSAVFNAD-----PEYTRQYETGVKSSWLDRLSTLSAYQIERFN 540

QY 562 AAPLPNNKTRYAALGKRVMEGVETEISGAVTPKWQIHAGYSYLHSQIKTASNRDD-- 620

Db 541 IRYRPDPKNPNFYAVSGKHSRSGVELSAIQIIPK-KLYLRGSLGVQAKVVEDKKNPDR 599

QY 621 --GIFLLMPKH-SANLWTTTYQVTPBELTIGGVNAMSGITSSAGMHAG-----G 665

Db 600 RVGIHLNNTSNVTGNLFRRY--TPTENLYGEI---GVGTG-KGRYGYDSRNKEVTTLPG 652

QY 666 YATFDAMAAVRFPTKCLKLOINADNIFNRHYVYRVGGANTFNIPGSERTWTANLYSF 722

Db 653 FARVDAMLGNWH-KNVNVTFAANLFNQKW-----RSDSMFGNPRGYTARVNYRF 702

RESULT 24

```

AAB68919
ID AAB68919 standard; Protein; 703 AA.
XX
AC AAB68919;
XX
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis protein #18.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dba; fhaB; fhuA;

```


KM Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KM immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KM systemic infection; non-diarhoeal infection; septicemia;
 KM pyelonephritis; antibiotic resistance.

OS Escherichia coli.

XX WO200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP03445.

XX 10-MAR-2001; 2000FR-0003145.

XX 02-FEB-2001; 2001FR-0001449.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Bingen E, Bonaccorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.

PT A library of DNA fragments of Escherichia coli strains for the
 PT phylogenetic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A- -

XX Example 6; Fig 6; 646p; English.

XX The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
 CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.

XX Sequence 721 AA;

Query Match 8.3%; Score 313.5; DB 22; Length 721;

Best Local Similarity 22.5%; Pred. No. 3.4e-19;

Matches 171; Conservative 127; Mismatches 344; Indels 119; Gaps 33;

QY 9 RINMTAATVLAALSSV---FAAQTADLETTHIKQORSYNALVTEKNGDYSSPAVTVGT 64

DB 2 KQITFAALLPALASFIPLHAASSTSEDEMI-VTGNTAADTTDSAAGAFKTNIDIVGP 60

QY 65 KIPASAREIPQSVITTNQOVKRVNDTPQARKTPGLRVLSN---DGRSSVYARGYE 121

DB 61 LGTSMIEFPYSTVTKEKIEHQOASVSEMKYSPSTOMARGGWDVGRPO--SRGMQ 118

QY 122 VS---EYNIDGL-----PAQMOSINGTLPULPAFDRVEMRGPGLDSDSGEGGIV 170

DB 119 GSVVANSRLDGLNIVTAFVEM-----LERNDVLNSLTGALYGPASPAGOP 166

QY 171 NLVRKPTPAFGHAAAGFTTHQYKAADVSSLSNDSGVRGVAAQTVGASPPRAEKN 230

DB 167 NFVAKRPTEETLRKVTLYGQSRSAFGHADLGHFDEKNRFGVRVILLDDEGSGVNDST 226

QY 231 NHEETVAAADMPINPDYV---GAGYLY-QQHLLAPYNGLPADANKKLPSP---QHY 262

DB 227 LRKLIVSVALDMWNIQGTQQLDASHSEFLQKQVGSFNNGP--VVKLPASANPKDKXL 263

QY 283 FVGADNKKFEMNSHDVADLKHYFG-----NGGYKVKYKRSRSDADS-----NYAFA 330

DB 284 ALSTAQN--DLTDTITSTRLIHFNDMSNAG---VGMQOABRARSVSSTKLNNQGD 338

QY 331 GSKLGKMTKTPAG--RPGCNTADDKACAVGLGTEIKOKALAFDASYSRPFRLGNTANEVIG 368

DB 339 SRMKDSTAGRRFVLSNTA-----GLNGHIDTGSIGHDLSLTGYWLSYSAKGTG 391

QY 389 ADVNRRSTNEOGRFTLYARGGLALNEFRSIPQVDLIANARKGVRSHTVATENIDFEG 448

DB 392 SSYS-----WGTMMYTPD-AIDE-----QGD--GKIRFGPRYRSVNTQSVTIG 435

QY 449 IYKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKETGVAVAYDLNNSLYL 508

DB 436 --DVTFTFQWSAMFYLSQSWLQTKNYDKHGQNTQVNDENGSLSPNALMYKITPNTMAYV 493

QY 509 ----SLSQLYTQPT--NLDAQKLLKPRQGNQFEVYKSYMDRLNARVSFRPMQXA 562

DB 494 SYADSLQGGTAPDESIVKNAQOTLNPYRSKQYEVGLKSDI--GEMNLGAALFRL--EBP 549

QY 563 AAPLNPNKKTRPAALGKRVMEGETEISGAVTPKQIHAGYSYLSOIK-TASNSRDQ 621

DB 550 FAYLDTDNV---YKEQGNQVNNGLLTAAGNWOGLNITSVTFDLPKLDRTANASTSNK 606

QY 622 IFLLMPKHSANIMTYYQV--TPELTIGGCVN-----AMSGITSSAGMHAGVATPDAMA 674

DB 607 QYVGVFKVQANLAEYSLPSIPEWVYSAVHTGKRAANDTNTS---YASSTYTD--LG 661

QY 675 YRFTPKLK----LQINADNIFNRHYAARVGAANTFNIGS 710

DB 662 TRYTKVSNVPTFRVYVNVNVEDKHYMASIFPSGTDGNGS 702

RESULT 28

ABBS8918

ID AABBS8918 standard; Protein; 700 AA.

XX AABBS8918;

AC 18-APR-2001 (first entry)

DE Neisseria meningitidis protein #17.

XX Meningococcus; meningitis; bacteraemia; vaccine; dba; fhaB; fhuA;

XX rnf; rth; toLC.

XX Neisseria meningitidis.

OS EPI069133-A1.

PN 17-JAN-2001.

PF 13-JUL-1999; 99EP-0401764.

PR 13-JUL-1999; 99EP-0401764.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PA (PLC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PI Nassif X, Tinsley C;

XX WPI; 2001-082916/10.

DR N-PSDB; AAF56458.

XX Immunogenic polypeptides derived from Neisseria meningitidis and the

PT nucleic acids that encode them, useful for diagnosing and vaccinating

PT against Neisseria infections e.g. bacteraemia and meningitis -

XX Claim 3; Fig 17B; 240p; English.

XX The present invention provides the protein and coding sequences of
 CC several genes from Neisseria meningitidis. These include the dba, fhaB,
 CC fhuA, rnf, rth17, rth19, rth20, rth21 and toLC genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.

XX Sequence 700 AA;

```

Query Match      8.1%; Score 307.5; DB 22; Length 700;
Best Local Similarity 21.9%; Pred. No. 1.1e-18;
Matches 138; Conservative 109; Mismatches 285; Indels 99; Gaps 24;

QY 8 FRINTAATVLAALSSVFAAOTAD-----LETVHIKQORSYNAIVTEKNGDYSSFA 59
DB 5 FHLALLPTLIIA--SPVAAADTDQNGEHYATLPTVSVVGSDTSVLKGYIN--YDEAA 60
QY 60 VTVGTRKIPASLREIPOSVSIITNQVKDRNVDTFDQARKTFLGLRVLNSDDGRSSVYARG 119
DB 61 VTRNGOL----IKETPTIDTLNIQKNKNGVNTDLSILLEGNAIDAAYDMRG-ESIFLRG 116
QY 120 YEYSEYNI--DGLPAQMQSINGTLNLFADFEVVRMGPSGLFDSGEMGGIVNLVRKP 177
DB 117 FOADASDIYRDGVRESGO--VRRSTANI---ERVEILKGPSSLYLGRGTGGGVINMVSKYA 172
QY 178 TKAFQGHAAAGCTHKQYKAEADVSGSLNSDGSVR-----GR-----VMAQTVGASPR 225
DB 173 NFKQSRNIGAVYGSRRANRSLNMDINEVLNKNVAIRLTGEVGRANFRSGIDSKNVWVSPS 232
QY 226 PAEKNNRHETFYAAADWDINPDTVLGAGLYQOORHLAPYNGLPADANNKLPSPQHVFG 285
DB 233 ITVKLDN-----GLKWT-----GQTYDNVERTPDRSPTKSYDRF-GLPYRMGF- 276
QY 286 ADWNPKNMNHVDF-ADLKHYFGNGYGVKGMRYSDRDRDADSNYAFAGSKLGMKMTAGRPG 344
DB 277 AHRNDFVKDKLQVRSRDLEYAFNDKWRAQMLAHRTAQDFDHTAGSENGNLIKRNYAW 336
QY 345 CNTADDKACAVGLGTEIKALAFDASYRPFRLGNTANEFVIGADYRFRSTNEQGRIT 404
DB 337 QQT-----DNKLTSSNFTLNGDYTIQGFENHLTGVMDYSK-----EHRNPT 377
QY 405 L-YARGGLALNEFRSTPQVDLIANARKVGRGYSHTVATEN---LDEFGYKSTFHPADG 460
DB 378 LGYSRAFTA-----SIDPYDRASWPASGRQLPILIQNRHKADSYGIFVQNIISATPD 429
QY 461 LSLIGGRGLGHYKISGECKTLHKASKTKFTGYA-----GAVYDLNDDNLSLSQLYT 515
DB 430 LKFLVGLGRYDKYTFNSENKLT---GSSROYSGHSPNIGAVWNNIPVHTLYASYNKGA 486
QY 516 P-----QTNLDADGKLL---KPRGNOPEVGYKGYMDRLNARVSVFVKDKKAAAPL 566
DB 487 PYGGGGYLSIDTSSSAVFNADPEYTRQYETGVKSWLNDRLDITLSAYQIBERFNRYP 546
QY 567 NPNKTRRYAALGKRVMEGVETEISGAVTPK 597
DB 547 DAENNPYTWAVGGKHSRSGVELSALGQIIPK 577

RESULT 29
AAR75366
ID AAR75366 standard; Protein; 696 AA.
XX
AC AAR75366;
XX
DT 13-MAY-1996 (first entry)
XX
DE Adhesin.
XX
KW adhesin; plasmid pear; vector; vaccine; intestine colonisation.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Misc-difference 293
FT /note= "deduced residue from nucleotide sequence
FT is His, this differs from Seq ID5, residue
FT is Gly"
XX
PN W09600233-A1.
XX
PD 04-JAN-1996.

```

```

XX 07-JUN-1995; 95WO-US06994.
XX 24-JUN-1994; 94US-0265714.
XX (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
XX (UNIW ) UNIV WASHINGTON.
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX Besser TE, Bilge SS, Tarr PI, Vary JC;
XX WPI; 1996-068826/07.
XX N-PSDB; AAT10105.
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
XX isolated on plasmid pSC (overlap), for use as a vaccine to prevent
XX bacterial colonisation of bovine intestine
XX Disclosure; Page 31-32; 42pp; English.
XX A candidate adhesin (AAR75366) was identified that is a homologue
XX of the IrgA protein of Vibrio cholerae. The adhesin enables
XX Escherichia coli O157:H7, an antibiotic-resistant, virulent and
XX common food-borne pathogen, to adhere to epithelial cells.
XX Recombinant adhesin was obt'd. by expression of the encoding sequence
XX (see AAT10105) in E. coli HB101 (pear). The adhesin can be used as a
XX vaccine for immunisation of cattle against disease or colonisation
XX of mucosal surfaces by O157:H7, thus increasing the safety of food
XX derived from cattle.
XX SQ Sequence 696 AA;
XX
Query Match      7.8%; Score 294.5; DB 17; Length 696;
Best Local Similarity 20.7%; Pred. No. 1.7e-17;
Matches 172; Conservative 103; Mismatches 302; Indels 255; Gaps 37;

QY 9 RINTAATVLAALSSVFAAOTADLETVHIKQORSYNAIVTEKNGDYSSPAVTGTKIPA 68
DB 2 RITTLASVWIPCLGFS--ASSIAAEDVMIVSASGY-----EK----- 37
QY 69 SLREIPOSVSIITNQVKDRNVDTFDQARKTFLGLRVLNS--NDDGRSSVYARG--EYSE 124
DB 38 KLTWAASVSVISQEEQLSSQYHDLAALRSVEGVDSGTGKGLGISIRGNPASYTL 97
QY 125 YNIDGL-----PAQMOSIN-GTLPNLFADFEVVRMGPSGLFDSGEMGGIVNLVR 174
DB 98 ILIDGVROGGSSDVTPTNGFSAMNTGFMPPLAAIERIEVIRGPMSTLYGSDAMGGVNIIT 157
QY 175 KRPTKATQGHAAAG-----FGTHQYK-----AEADVSGSLNSDGSVRGR-----V 215
DB 158 RKNADKWLSSVYNAGLNLSQSNKNGNSQFNFWSSGPLVDDSVLSQVRGSTQORQSSVTS 217
QY 216 MAQTVGAS--PRPAEKNNRHETFYAAADWDINPDTVLGAGLYQOORHLAPYNGLPADANNK 274
DB 218 LSDTAGTRIPYPTESQYVN--LGLRLDWKASEQDVL----- 251
QY 275 LPSLPQHFVFGADWNKFKMNSHDVFDLKHFGNGGYGVKGMRYSDRDRDADSNYAFAGSKL 334
DB 252 -----W--FDMDT-----TRQRYDNRDGO-----LGSALT 273
QY 335 GMKTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYRPFRLGNTANEFVIGADYRNR 394
DB 274 G-----GYDRTLRYERNKISAGYDGTFTFTGWS-----YLNW 306
QY 395 RSTNEOGRITLYA-----RGGLALNEFRSIPQVDLIANA-RKGVGRGYSHTV----- 440
DB 307 NETENKRELVRSVLKDKWGLA-GQPRELKNLSNLSLLTLPGLSHLTVTCGEGFQSS 365
QY 441 -----TENLDEFGIYKSTFHPADGLSLICGGRLGHYKIESGEGKTLHKASK 487
DB 366 SMKDGVVLASTGETFRQKSWSPAEDEWHLTDLALTAGSYEHHEHGFQGH----- 416
QY 488 TKFTGYAGAVYDLNDDNLSL-----YLS--LSQLYTPQTNLDADGKL-----LKPRQ 531

```



```

Db      417 --FSPRAYLVMDADADMTLGGVTTGKAKRMQQLHKGISGVGQGTNLLGNPDLKPEE 474
Qy      532 GNQFVEYKGSYMD--RLNARVSFYRMKDKNAAAPLNNKTRVYALGKRVMEGYETE 589
Db      475 SVSEYAGV---YYDNAGALNANTGFWTDSNNKIVSINDNNTSYNSGKALHGV--E 529
Qy      590 ISGAVTPKQO---IHAGSYLHSQIKTASNSRDDGILFILMPHSANLMTTYQVTPLELT 645
Db      530 FAGTL-PLWSEVDYTLISLNTYWTSEQRDGNK--GAPLSTYPEHMYAKLNMQITEVAS 586
Qy      646 GCGV-----NMSGITSSA-----GMHAGYATFDMAAVRFTPKLQINAD 688
Db      587 WLGARIRGKTPRTQNVSSLSAVQKKYDEKGYLKAWTVVDAGLSKMTDALTLNAAVN 646
Qy      689 NIFNRHY-----YARVGANT-FNIPGSEPTWTANLARYSF 722
Db      647 NLNKKYSDVSLXSAGKSTLYAGDIPQTSSTTGVIYF--ERYWMSLNTQF 696

```

RESULT 30

AA72914
ID AA72914 standard; Protein; 725 AA.

AA72914;

13-JUN-2001 (first entry)

E. coli ironNc extracytoplasmic protein fragment.

ironNc; extracytoplasmic protein; immunogen; vaccine; E; UTI;

immunotherapy; extraintestinal infection; urinary tract infection;

meningitis; pneumonia; intra-abdominal infection; antibiotic.

Escherichia coli.

Key Location/Qualifiers

Peptide 1..24 /label= Signal_peptide

Protein 25..725 /label= Mature_E. coli_ironNc_extracytoplasmic_protein

Misc-difference 120..121 /note= "Encoded by GTG CGT TAT"

Misc-difference 134..136 /note= "Encoded by AAC TGG"

MO200121636-A1.

29-MAR-2001.

22-SEP-2000; 2000WO-US26117.

22-SEP-1999; 99US-0155621.

(UTM) UNIV NEW YORK STATE RES FOUND.

Russo T, Carliano U;

WPI: 2001-244936/25.

N-PSDB; AAD03054.

Novel isolated ironNc polynucleotide from extraintestinal isolate of

Escherichia coli useful as vaccine for treating or preventing

extraintestinal infections caused by extraintestinal pathogenic

Escherichia coli -

Claim 9, Page 39-41; 44p; English.

The invention relates to the identification of ironNc gene,

from an extraintestinal isolate of Escherichia coli. This

gene is expressed in increased amounts in human urine and is

identified by transposon (TnpA) mutagenesis. ironNc gene

encodes an extracytoplasmic protein. This gene can be used as

an immunogen in vaccine formulations. The recombinant vector comprising nucleotide sequence encoding one or more antigenic epitope of ironNc is useful for diagnostic and immunotherapeutic purposes. The ironNc antigenic peptide is useful for treating or preventing extraintestinal infections (EIs) caused by extra-intestinal pathogenic E. coli (ExPEC). The EIs include urinary tract infection (UTI), meningitis, intra-abdominal infection and pneumonia.

The present sequence is E. coli ironNc extracytoplasmic protein.

Sequence 725 AA, Query Match 7.2%; Score 273.5; DB 22; Length 725;

Best Local Similarity 19.8%; Pred. No. 1.5e-15; Matches 167; Conservative 127; Mismatches 304; Indels 247; Gaps 40;

5 MSYRINMTAAATVLAASSVPAQCPADLETHIKQSRNALVTEKNDYSFAVTGT 64

1 MRINKLMTLTVLVGLNQSVAKYS-----DDNDETLVVA 39

65 KIPASLRRIPOSVITITNOQVDR--NVDTFDQARKTPGLRVLSN-----DDGRSSVY 116

40 TAEQVLKQCP-GVSVITSEDIKTTPVNDLSDIIRMPGVNLTGNSASGTRGNRQIDR 98

117 ARGVEYSEYNIDGLPAQMS-----INGTLPLNFAPDVEMRGPSGLPDS 162

99 GMGPENTLILIDGVPVTSRNSVYSWRGERDTRGDTRWVP--EQVERIEVIRGPAARVG 157

163 SGEMGIVNLVLRKPTKAAQGHAAAGFTHKQYKAA-----DVSGLNSDG--SV 211

158 SGAAAGVAVNIITKRPNDV--HGSLSLYTNQPESSSEGAIRANFSLSPGLADALTTBL 215

212 RGRV-----MAQTVASPPRAEKNNRHETFYAAAMDINPDVLG--AGLYQQRH 260

216 YGNLAKTIDADSMNDINSPVGTAKAAGHEGVRNMDINQVMSKLNPOQIILPEVGYSRQ-- 273

261 LAFYNGLPADANNKLPSPQWV--GADMNKFKMNSHDVFDLKYFPNG--GYGRV 314

274 ---NIVAGDTQSSSSAVTESLAKSGKETNRLYRQNGI---TH--NGIWDQGSRF 322

315 GMRYSRDLSDNVAFAAGSLGKMTKPRGRCNTADDKAAVGLGTIKOKALAFDASYSR 374

323 GYVEKTN-----NTMNBGLSGGSGRILLAGK-----FTNRLS 358

375 PPLIGNTAN-----EFVIGADYNRFRSTNEQRTTLVARGGLALNFRSIPQVDI 425

359 SWRTSGELNIPLVNMTDQTLTVGAEMNRDKLDPSST--LTVND-RDI----- 404

426 ANARKGVGYSHVATENLDEF-GIVGKSTFPADGLSLIGGR----- 468

405 ---SGISGSADRSSKNSQISALYIEDNIEVPETNIIIGLRFYLDSDGSGFSPSLN 460

469 ---LGHY-KISGSGKT-----LHKAATKFTGVA-----GAYDLDNNSLYLSLQ 512

461 LSGELDYFKVKAQVARTKAPNLQSS--GYLVKSGNGCPDINSGCCYLLG--- 513

513 LYTPTNLADGKLLKPRQNGPEVYKGSYMDRLNARVSFYRMKDKNAAAPLNNK 572

514 ---NKLD--PEISVNKEIGLETFWED--YHSAVTFIRNDYOKKIYAGDVITQ 560

573 TRYAL-----GKRVMEGVETEISGAVTPK---WQHAGSYLHSQIKTASNSRDDGI 622

561 TSGAVILKMQNGKALVDIGIEASMFPLVKRLMNTATM-----MISSEQDTGN 613

623 FL-LMPKHSAN-----LWTTY-----QVTPELTTGGVNVANSGITTS 657

614 PLSVIPKYTIINSLMTITQAFSASFMTWLYGRQKRTHTAETRSSEPT--GGL----- 663

658 SAGMHAGVATPDMAAAYFTPKLKQIINADNIPNHYAVARQANTFNIPOSERTWTAN 717

664 -SGKEIGAVSLVGTNPNYINKNLRLNVGVSINLKKQIRSRSEGANTYEPG--RAYVAG 720

Qy 718 LRYSF 722

Db 721 VTASF 725

RESULT 31

AA68924
ID AAB68924 standard; Protein; 697 AA.

AC AAB68924;

DT 18-APR-2001 (first entry)

DE Neisseria meningitidis protein #23.

XX Meningococcus; meningitis; bacteraemia; vaccine; debA; fhaB; fhuA;

KW rni5; rth; to1c.

OS Neisseria meningitidis.

XX EP1069133-A1.

PN 17-JAN-2001.

PF 13-JUL-1999; 99EP-0401764.

PR 13-JUL-1999; 99EP-0401764.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PI Nassif X, Tinsley C;

DR WPI; 2001-082916/10.

XX N-PSDB; AAP56464.

DR Immunogenic polypeptides derived from Neisseria meningitidis and the

PT nucleic acids that encode them, useful for diagnosing and vaccinating

PT against Neisseria infections e.g. bacteraemia and meningitis -

XX Claim 3; Fig 23B; 240pp; English.

XX The present invention provides the protein and coding sequences of

CC several genes from Neisseria meningitidis. These include the dbaA, fhaB,

CC fhuA, rni5, rth17, rth18, rth19, rth20, rth21 and to1c genes. These can

CC be used in the diagnosis and treatment of infection by the bacterium,

CC which can lead to meningitis and bacteraemia, and in vaccines to prevent

CC such infection.

XX SQ Sequence 697 AA;

Query Match 6.4%; Score 241.5; DB 22; Length 697;

Best Local Similarity 20.8%; Pred. No. 1.2e-12;

Matches 126; Conservative 99; Mismatches 245; Indels 135; Gaps 25;

Oy 8 FRINMTATVLAALSSVFAQTAD-----LETVHKQGRSYNAIVTEKNGDYSSFA 59

Db 5 FHLALLPTLIIA--SFPVAAADTDQNGEHYATLPTSVVVGSDTSVLKGYIN--YDEAA 60

Oy 60 VTVGTKIPASLREIPQSVSIITNQVQKDRNVDTFDQARKTFLRLVLSNDDGRSSVYARG 119

Db 61 VTRNGOL----IKETPTIITLNIQKNKNGVGTNDLSILEGNAGIDAAVDMRG-ESIFLRG 116

Oy 120 YEYSEYNI--DGLPAQMOSINGTLNPLNFAFDRVEMRGPGLFDSGEMGGIVNLVRKRP 177

Db 117 FQADASDIVRDGVRESGQ-VRESTANI---ERVEILKGPSSVLYKRTNGGGVINVMSKYA 172

Oy 178 TKAFOGAAGAAGCTHKQYKAEADVSGSLNSDGSVR-----GR-----VMAQTGASPR 225

Db 173 NFKQSRNIGAVYGSWANRSLNNDINEVLNKNVAIRLTGEVGRANSFRGIDSKNMVWSPS 232

Oy 226 PAEKNNRHETFYAAADWDINDPTVLGAGLYQORHLP-----YN--GLPADANNKL 275

Db 233 ITVKLDN-----GLKWT-----GQTYDYNVERTPDRSTKSVYDRFGLPYRWGFAH 278

Oy 276 PSLPQHVFVGADWKNFKMNSHDVF-ADLKHYFGNGGKVGKGMRYSDRDADSNYAFAGSKL 334

Db 279 P-----NDFVKDKLQVWRSDLEYAFNDKWRQAOMQLAHRTAQDFDHFYAGSEN 326

Oy 335 GMKTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRF 394

Db 327 GSRIKRNVAWQOT-----DNKTLSSNFTLNGDYITIGRFENHLTVGMGYSR- 371

Oy 395 RSTNEQGRITLYARGGLALNEFRSIPQVDLIANARKVGRGYSHTVATEN---LDEFGIYG 451

Db 372 ----EHRNPTLIGRGSP-----TVP-INPYDRASWPASGRQLQPILTQNRHKADSYGIFV 420

Oy 452 KSTTHPADGLSLICGGRLGHYKIEGEGKTLHKASKTKFTGYA-----GAVYDLNDNNSL 506

Db 421 QNIFSATPDLKFLVIGRGYDKYTFENSENKL---GNSRQYSGHSPNIGAVWVNNIPVHTL 477

Oy 507 YLSLSOLYTPQTNLDADGKLLKPRQGNQFVGKYGKSYGMDRLNARVSYFRMKDKNAAPL 566

Db 478 YASYNKGFAP-----YGGRGYLSIDTSSAAVF-----NAAPEY 510

Oy 567 NPNNK 571

Db 511 TPNTK 515

RESULT 32

AAU33589

ID AAU33589 standard; Protein; 813 AA.

AC AAU33589;

XX 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #33.

XX Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX Pseudomonas aeruginosa.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS51448.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 5085; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acid can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX
SQ Sequence 813 AA;

Query Match 6.3%; Score 237.5; DB 22; Length 813;
Best Local Similarity 20.6%; Pred. No. 3.5e-12;
Matches 171; Conservative 107; Mismatches 301; Indels 253; Gaps 38;

QY 19 AALSSVFAAQTADLETTHIKGQSRNATVTEKNGDYSSPAVTVGKIP-ASLRRIPO-S 76
DB 107 AAVPATOASEPLDHEQMETTSRTSSDLVS-----ATROSTVIEHAQLEEDRQS 157
QY 77 VSIITNQYKDRNVDTFDQIARKEPGLRVLSNDGRSSVYARGYESE-----YNID 128
DB 158 DSLAT-----VLAKVPGM-----SDSSKRTIT-----EYGQTLRGSMVLWMD 195
QY 129 GLPAQM-OSINGTLPNL--PADRVVVMKPGSGLEPDSSEMGIVNLVKRPKFAQGH 185
DB 196 GYPLMTNRDSSNNLANIDALIERIVIRGSSAIY-GSGATGGIIS-ITTRP----- 245
QY 166 AAGFGTHKQYKAADVSGS-----LNSDGSVSG--RWAAQTGASPRRAEKNRHETV 237
DB 246 AGG-----ENRAETRLSATSPYLRIGSDG-LGQFOQYFAGSGALUDYSPDFGTRH----- 295
QY 238 AAADWDINPDYVL--GAGVLYQQRHLAPYN-----GLPADANN----- 273
DB 296 VQASVDHAGDRILAPESQGDLPDSN--VYNIGKGLAIDENQROVALSHVDARQDPT 352
QY 274 -----KLP--SLFOHVFVQADNMKFKMNSHDVADLKHYFNGGQYGVKGRYSDRDA 323
DB 353 YATDPRVALPQGSVYANAIKGLELDE-QNRIRNTLANLEY----- 392
QY 324 DSNVAFAGSKLGMK-----TPAGRPGCTADADKACAVGLGEITKOKALAPDASY 373
DB 393 -ENLDLIGRLSAQLYYRDYFTRFTFPDARAVSTRG-----GNVDQIMQNSEVFGSRLT 445
QY 374 RPPRLGNTAN-EFVIGADYNRFRS--TNEQGRITLYARGGLALNFRSIPQVDLIANAR 429
DB 446 LATPLGESGNTLVWGGDINQGRSDMPDLVFPDAADAGGLVFDKIGLITTPMLPRTS 505
QY 430 KCV-----RGYSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKI 474
DB 506 AGAFPAQLQHRFDEHWSIDGLGLEYESTAEFDDEPIPLSES--KAASPVTVKGGDLIDYAN 562
QY 475 EGGEGKTLHAKSKTKRTGYAGAVYDLNDNNSLYLSL-----QYTPQTNIDAG 524
DB 563 LSN-----LGIVYSPVACQEIYASFSQGFOLDVGIQLRNARGFIDGS 606
QY 525 KLLKPRQGNQFEVGYKGYMDRLNARVSFYRMKDKNAAPLIPNNKKTIRYALGKRV 584
DB 607 SNIPEPKTNNYELGMRGALCGNTLSGLAFYTTSSKLGDVQSPFNGLILR-----TKERY 662
QY 585 GVE-----TEISGAVTP--KWOIHAGY-----SYLHSQIKTAS 615
DB 663 GVEASADWLSDDEVMGAGSATWNRGRERKPDGKDMQGVYRVPPLKLAIVLYCKRDADW 722
QY 616 NSRDSOIFILMPRHGSANMTTYQVTEPLTIGGVNAMSQITSSAGNHAGYATFDMAAY 675
DB 723 NNRLQATFPDSDKY-----RLDGVESFRQVOSTYTTTVDLVSOY 761

QY 676 RPTPKIKIQINADNIFNRHYARV-----GGANTFNIPGSEPTWTANLRYSF 722
DB 762 RTTPDQSLIGIQLFNFRDYPLYSQLLNNNNNTSHLPAPGTVLASYYTHNW 813

RESULT 33

ID AAY38830 standard; Protein; 211 AA.

AC AAY38830;

DT 08-OCT-1999 (first entry)

DE Neisseria meningitidis antigen encoded by a partial ORF23.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
treatment; Neisseria infection; meningitis; septicemia; gonorrhea.

OS Neisseria meningitidis.

PN W0924578-A2.

PD 20-MAY-1999.

PF 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

PA (CHIR-) CHIRON SPA.

PI Grandi G, Maignant V, Pizza M, Rapuoli R, Scarlato V;

DR WPI: 1999-327407/27.

DR N-PDB: AA12258.

XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for

PT diagnosis, treatment and prevention of infection

PS Claim 4; Page 379; 524pp; English.

XX Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*

CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open

CC reading frames (ORFs) AA21972-212358. The antigenic proteins,

CC their fragments, their nucleic acids and antibodies are used for

CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*

CC infections, such as meningitis, septicemia and gonorrhea. Both

CC organisms are closely related. Fragments of the nucleic acids

CC are useful as hybridisation probes and antisense reagents.

XX
SQ Sequence 211 AA;

Query Match 6.2%; Score 235.5; DB 20; Length 211;
Best Local Similarity 31.5%; Pred. No. 6.2e-13;
Matches 64; Conservative 34; Mismatches 94; Indels 11; Gaps 3;

QY 111 GRSSVYARGYEVSEYNIDGLPAQMOSINGTLPNLPFAFDRVEYVMRGPGLFDSGEMGIV 170

DB 1 GYNYLFAKRSRIANQIINIPVADALADGANTAYEVEYVRYGAGLLDGTGSESATV 60

QY 171 NVRRKPTKAFQGHAAAGFTKQYKAADVSGSINSDSVGRVWAQTVGASPRPAEK 230

DB 61 NVVRKRLTRKPLFEVRAEAGNKRHFGLDADVGSINTEXXLRGLVSTGRGDSWRERR 120

QY 231 NNHETFYAADDINDIPDYVLGAGIYQQRHL--APYNGLPADANKPLSLQHFVVGAD 287

DB 121 SRXAEIYGLIEVDIAFQFRVHAXMDYQQAQKERTADAPLSYAVVDSQYATAFGPKDNPATN 180

QY 288 WNKFKMNSH-----DVPADLKHYF 306
 Db 181 W-----ANSHRALNLFAGIEHRF 199

RESULT 34

ABBS2669
 ID ABBS2669 standard; Protein; 660 AA.

XX AC ABBS2669;
 XX DT 11-FEB-2002 (first entry)
 XX DE Escherichia coli polypeptide SEQ ID NO 732.

XX KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicemia;
 KW pyelonephritis; antibiotic resistance.

XX OS Escherichia coli.

XX PN WO200166572-A2.

XX PD 13-SEP-2001.

XX PF 12-MAR-2001; 2001WO-EP03445.

XX PR 10-MAR-2000; 2000FR-0003145.

XX PR 02-FEB-2001; 2001FR-0001449.

XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX DR WPI; 2001-550253/61.

XX PT A library of DNA fragments of Escherichia coli strains for the
 PT phylogenetic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A-
 XX Example 6; Fig 6; 646pp; English.

XX CC The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.

XX SQ Sequence 660 AA;

Query Match 6.0%; Score 228; DB 22; Length 660;

Best Local Similarity 19.7%; Pred. No. 1.8e-11;

Matches 157; Conservative 107; Mismatches 276; Indels 258; Gaps 37;

QY 3 QFWSVRINMTAATVLAALSSVFAAQTADLETVHIKQORSNAIVTEKNGDYSSFAVTV 62

Db 5 QFTSLRLSLLALAVSATLPTAFATET-----MTVTATGNARSSF----- 44

QY 63 GKIPASLREIPQSVSIITNOQKDRNVDTFDQLARKTFLGLRVLSNDDGRSS---VYARG 119

Db 45 -----EAPMVMYVIDTSAPENQATATSDLLRHVPGITL--DGTGRTNGQDVMNRG 93

QY 120 YEYSE--YNIDGLPAQMOS--INGTLNLFADFDRVEVVRGPGSLFDSSGEMGV----- 170

Db 94 YDRGVLVLVDGVRQGTDTGHLNGTFLDPLALKRVEIVRPSALLYGSGALGGVVISYDTV 153
 QY 171 ---NLVRRKPTKAPQGHAAAGFGTHKQYKABADYSGSLNSDGSV-----RGRVMAQTVG 221
 Db 154 DAKDLLQEGQSSGFRVFGTGTGDHSLGLGASAFGRTEINLDGIWAMSSRDGDLRQSNGE 213
 QY 222 ASPPRAEKNNRHETFYAAADWDI-----NPDVLGAGLYYQQRH 260
 Db 214 TAPNDESINN---MLAKGTWQIDSAQSLGVLRYNNDAAREPKNPQTVEASD---SSNP 266
 QY 261 LAPYNGLPADANNKLPSPHQHVFVGADWNKFKMNSHDVFADLKHYFGNGVKGMRYSD 320
 Db 267 MVDRTTQORDAQLSYKLAPO---GNDW-----LNADAKY-----WSEVRINQON 308
 QY 321 RDADSNYAFAGSLGKMTKTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPFRLGN 380
 Db 309 TGSSEGREQITK-----GARNRSTLTFADSF----- 337
 QY 381 TANEFVIGADYNRRSTNEQRTTLYARGGLALNE-----FRSIPQVDLIANAR-KG 431
 Db 338 -SHLLTYGGEYR-QEQHPGGATTGFPQAKIDFSSGMLQDEITLRDLP-ITLLGGTRYDS 394
 QY 432 VRGYSHTVATENLDEFGIYKSTFHPADGSLGGRGLGHYKIESGEGKTLHKASKTKFT 491
 Db 395 YRGSSDGKVDADKSSRAGMTINPTNWLMLFG----- 428
 QY 492 GYA-----GAVYDLNNDNNSLYLSLSQLYT---POTNLADADGKLLKPRQGNQFEVGY 539
 Db 429 SYAQAFRAPTMGEMV---NDSKHFSIGRFYTNVWPNP-----LRPETNETQEYGF 477
 QY 540 KGSYMD-----DRLNARVSFYRMKDKN-----AAAPLN---PNNKTRYAALGKRV 582
 Db 478 GLRFDDLLMSNDALFPAKYISTVDFAAATTMSYVNPNAK-----IWGWDV 532
 QY 583 MEGVETEISGAVTPKWOIHAGYSYLHSQIKTASNSRDDGIFLLMPKHSANLWTTVYQVTP 642
 Db 533 MKTKTTDL-----FSLDVAYN-----RTRGKDDTGTGEYI-----SSINPD 567
 QY 643 LTIGGGVN---AMSGIT-----SSAGMHAGGYATFDAMAAYRFTPKLK--- 682
 Db 568 -TVSTLNIPAHSGFSVGWGTFADRSTHSSYSKPGYGVNDFFVYSYQGGQALKGMT 626
 QY 683 LQINADNIFNRHYARVG 700
 Db 627 TTLVLGNADFKEYWSPQG 644
 AC ABBS2836;
 XX DT 11-FEB-2002 (first entry)
 XX DE Escherichia coli polypeptide SEQ ID NO 1077.
 XX KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 XX systemic infection; non-diarrhoeal infection; septicemia;
 XX pyelonephritis; antibiotic resistance.
 XX OS Escherichia coli.
 XX WO200166572-A2.
 XX PD 13-SEP-2001.
 XX PF 12-MAR-2001; 2001WO-EP03445.
 XX PR 10-MAR-2000; 2000FR-0003145.
 XX PR 02-FEB-2001; 2001FR-0001449.

PA (INRM) INSERT INST NAT SANTE & RECH MEDICALE.
 PI Bingen E, Bonaccorsi S, Clermont O, Nassif X, Tinsley C;
 XX WPI; 2001-550253/61.
 XX
 XX A library of DNA fragments of Escherichia coli strains for the
 PT phylogenetic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A- -
 XX
 XX Example 6; Fig 6; 646pp; English.
 XX
 CC The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
 CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.
 CC
 CC Sequence 682 AA:
 SQ
 Query Match 5 9%; Score 222.5; DB 22; Length 682;
 Best Local Similarity 20.1%; Pred. No. 6.2e-11;
 Matches 159; Conservative 116; Mismatches 313; Indels 203; Gaps 33;
 QY 15 ATVLAALSSVFAAQTADLETVHIKGRSNAIVTEKNGDYSFAVTVGKIPASIREIP 74
 DB 12 AVMCAVIVSSGYASSPKKEDTL-----VVTASGTFQOIRNAP 48
 QY 75 QSVSITITNOQVKRNVDTFDQARKTPGLRLVLSNDGRSSVYARGY--EYSEYNIDGLPA 132
 DB 49 ASVSIVTSEQLQKPYSDLDVAVKQDVEGISITGGE--KPDISIRGLSGDVTLLVVG--- 104
 QY 133 QMOGIN-----GTLPLPAFDRVEVWRGSPSGLPDSSGEMGIVNLVKKRPTA 180
 DB 105 RROSGRESRPNGSGGEAFPIPPVEAIEIRIEVIRGPMSSLYGSDAIGVYINITTK-PVNN 163
 QY 181 FQGHAAAGFG---THKOYKAADVSGSLNSDGSVGRVMAQGVASPPRAEKNNNHET 236
 DB 164 QTWGDTVLGIGIIOEHGF-----GNSITNDPFLSGPLIKDLGIQ----- 204
 QY 237 YAAADWDINDPVLGAGVLYOQRHLAPYNGLPADANNKLPSELPQHVFGADWNKFRPNSH 296
 DB 205 -----LVGGMVYRKED-SISQGTAPKDKNKITATIQ--FTPTESQK----- 243
 QY 297 DVFADLKHIFGNGYGVKGMRYSDRADSNVAFASGKLGMKTP-AGRPCCNTADDAACAV 355
 DB 244 -FEE-----YKKNQVHTLTPGESLIDMT-MRGVLMKQPSKRETHNSKSHVAAW 291
 QY 356 GLGTEI-----KOKALAPDASY-SRPRPLGNTANEPVIGADYNRF 394
 DB 299 NAGGELHEIAVYQEKVIREVSKGKKDKXNMWLVESKREPIYNT---IIDAVYTF 347
 QY 395 RSTNEGRTLLYARGLALNFRSIPQVDLIINARKGVGYSHTVATE--NLDEFGIYK 452
 DB 348 LPEN-----VLTIGG---QFOHAEIRDSATGK-----TBTQSVISIKQAVPIE 390
 QY 453 STFHPRADGLSLGGGRLGHYKTESGEGKTLHKASKTKFPGYAG---AYYDLNNSLYL 508
 DB 391 NEYATDLSLATLGGRLDNHEI-----YQSYNPRLYAYVYNNLDNLTLLKG 435
 QY 509 STSOLY-TPQTNLDADG-----KTLKPRQGNQPEVGYKSYMDRLANAV 552
 DB 436 GIAKAFRASTIEVSPFGFTLLVGGASIMYGNRDLKPEISVTEIIGIITS-NDSCFSASA 494
 QY 553 SFYRMKDKNAAPLNNNNKT-----RYAALGKRWMEGVETETISGAVTPKQIHAGYS 605

DB 495 TLENTDFKNKLTSTYDGTDPVTLNTFIYDVNGEANINGVELATQIPYDKRHVSANYT 554
 QY 606 YLHSLKTSANSRDDGIFP-----LMPKHSANLMTTYGVTELLTGGGVNAMSGITSSA 659
 DB 555 FTDSRRKSDDESL-NKSKLSKGEPLERTPRHANAUKLEWYTODITFSSLN-YTGQIWA 612
 QY 660 GWHAG-----GYATPDMAAAYRFTPKLKQINADNIFNHHYARVVGANTFNIPGSE 711
 DB 613 AQRNGAKVPRVANGFTSMIDIGLNTYQILPDTLLNFALVANTDKRSEDIDTIDGMWQDEGR 672
 QY 712 RTWTANLRYSF 722
 DB 673 RYW-ANVRVSF 682
 RESULT 36
 AA80096
 ID AA80096 standard; Protein; 785 AA.
 XX
 XX AA80096;
 AC
 XX 16-MAY-2000 (first entry)
 DT
 XX
 DE TonB dependent receptor fccL amino acid sequence SEQ ID NO:25.
 XX
 KM Bacterium 2412.1; maize; detoxification; degradation; carcinogen;
 KM fumonisin catabolic gene cluster; antifungal; mycotoxin.
 OS Bacterium 2412.1.
 PN WO200004158-A2.
 PD 27-JAN-2000.
 XX
 PF 14-JUL-1999; 99WO-US15837.
 XX
 PR 15-JUL-1998; 98US-0092953.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Duvick JP, Maddox J, Gilliam J;
 DR WPI: 2000-171270/15.
 DR N-PsDB; AA291265.
 XX
 PT New isolated nucleic acid molecules which encode fumonisin degradative
 PT enzymes and transporters, used for detoxifying fumonisin or related
 PT mycotoxin or reducing the pathogenicity of fungi producing fumonisin -
 PS Claim 1; Page 148-151; 181pp; English.
 XX
 CC The present invention describes isolated nucleic acid molecules which
 CC encode fumonisin degradative enzymes and transporters from a
 CC Bacterium 2412.1 fumonisin catabolic gene cluster. The Bacterium 2412.1
 CC was isolated from maize. The nucleic acid molecules can be used for
 CC producing an enzyme for detoxifying fumonisin or a structurally related
 CC mycotoxin, e.g. in harvested grain, or processed grain which is to be
 CC used as animal feed or silage. Ruminal microorganisms transformed with
 CC the nucleic acid molecules can be used as probiotic compositions or as
 CC feed inoculant compositions. Plants transformed with the nucleic acid
 CC molecules can degrade or transport fumonisin and can reduce the
 CC pathogenicity of a fungus producing fumonisin. The plants may be e.g.
 CC maize, sorghum, wheat, tomato, soybean, alfalfa, sunflower, Brassica,
 CC cotton or rice. The products can also be used as detection reagents for
 CC fumonisin and related compounds. AA291253 represents the Bacterium
 CC 2412.1 fumonisin catabolic gene cluster, and AA291254 to AA291273
 CC represent isolated nucleic acid molecules from Bacterium 2412.1 which encode
 CC AA80085 to AA80104, AA291286 to AA291291 represent sequence used in the
 CC exemplification of the present invention. AA291274 to AA291285 are given
 CC in the sequence listing from the present specification but are not
 XX specifically mentioned further.

SQ	Sequence	785 AA;
Query Match	5.8%; Score 220; DB 21; Length 785;	
Best Local Similarity	20.3%; Pred. No. 1.3e-10;	
Matches	162; Conservative 109; Mismatches 345; Indels 182; Gaps 32;	
Qy	9 RINMTAATVLAALSSVFA-AQTADLETVHIKQSYNAIVTEKNGDYSSFAVTVGTKIP 67	
Db	29 RLMLTAAGSALVGLAPKALQAVPPAGHEASQVEIIVTAQ-----RRS 75	
Qy	68 ASLREIPQSVSIITNQVKDRNVDFDOLARKTPGLRV-LSNDG-RGSVARGVEYSEY 125	
Db	76 ENIQNVPSVQALSAAQLEREGIKQTSDIARVTPNVTAMPNGEGNQPAVTIRGIGLNDP 135	
Qy	126 NIDGLPAQMOSINGTLPN-----LPAFDRVEVMRGPGLDFSSGEMG 167	
Db	136 N-----SNNAGPNAIYVDDVVISAPSQTFGIFDINQIVLKGPGQTLYGRNSSG 185	
Qy	168 GIVNLVRKPTKAFQGHAAAGFTHQKQKAEADVSGSL-----NSDGSVRGR 214	
Db	186 GALVFTSRAPSQDFAADAHFDYGSNTYQLQAGVGGLSDQLSARLAFVFNHSDGFMENT 245	
Qy	215 VMAQTVGASPRPAEK-----NNRHETFYAADWDINDPDTVLGAGYLYQORHLAPY-NG 266	
Db	246 LTGGSASGTDNOAVRLQLLYRNDRLKULLSSAYGHVNSPIV-----QYHGLGAPAG 298	
Qy	267 LPADANNKPLSPHQHVFGADWNKFMNSHDVFDLKHFGNGYGVKGMRYSD--RDAD 324	
Db	299 TQSSASPTLCS-PEQVRAGGCVNFGAGTPSGF-----YDSSDRGE-RLRVENFLQOAR 351	
Qy	325 SNY-----AFAGSKLGMKTPAGPGCGTADDKACAVGLGTEIKQKALAPDAYS 373	
Db	352 ADYEYGPVTLTISAFTHSK-----KSGPDDADGTSLSLLHATYGVRSDDTWTQEFRAAYS 406	
Qy	374 RPFRLGNTANEFVIGADYNRPRSTNEQGRITLY---ARGGLAL-----NEFRSIPQVDLIA 426	
Db	407 GO-RL-----HWVAGAYLDELTKQNQPLSIFYDGRFGGLIPARAGAFDGIAGKLSQ 460	
Qy	427 NARKGVGRYSHTVATENLDEFGI-----YKSTFHPADGLSLIGGRLGHY----XIE 475	
Db	461 NTQKTRSIAPFGQADYTLDRFTLTLLGGRYTHERKTFDHPSATQVQAGG-LGKYGPLGKIV 519	
Qy	476 S-----CEGKTLHKASKTKFTG--YAGAVVDLNDNSLYLSLSQL 513	
Db	520 SLSEAFKASDPTWRAALSYPRAVMYGVSVATPKGAGFNGGFLSSPNKAL-----572	
Qy	514 YTPQNLADGKLLKPRGNOQPEVGVKGSYMDRLNAR-VSPYRMKDK-----NAAAPLN 567	
Db	573 -----AAVPVAPEKVTTYELGFKSLPERLLVNGAAYFNSYDNEQILANTAVVD 624	
Qy	568 ----PNNKTRYAALGKRV-MEGVETEISGAVTPKWQIHAGYSYLHSQIKTASNSRDDGI 622	
Db	625 TVTGPVTVTVTLNARKAHSQGVLEKAVPIPLVLSLQPAWLRLTRLDEAGFSGGTSL 684	
Qy	623 ----FLMPKHSANLWTTVQVTPBELTIGGVNAMSGITSSAGHAGGYATFPDAAAYRFT 678	
Db	685 EKGQFHLADDDSVNV-----AFTSAYKSHQFPDSTNAPYTOQEGYVWVHNASLTFNSR 736	
Qy	679 PKLKLQINADNIFNRHY 696	
Db	737 NHWDVGFVNRNLGTQY 754	
RESULT 37		
AAV75536		
ID	AAV75536 standard; Protein; 188 AA.	
XX		
AC	AAV75536;	
XX		
DT	21-MAR-2000 (first entry)	
XX		
DE	Neisseria meningitidis ORF 743 protein sequence SEQ ID NO:2546.	

KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;	
XX	antibacterial; gene therapy.	
OS	Neisseria meningitidis.	
XX		
PN	WO9957280-A2.	
XX		
PD	11-NOV-1999.	
XX		
PF	30-APR-1999; 99WO-US09346.	
XX		
PR	01-MAY-1998; 98US-0083758.	
PR	31-JUL-1998; 98US-0094869.	
PR	02-SEP-1998; 98US-0098994.	
PR	02-SEP-1998; 98US-0099062.	
PR	09-OCT-1998; 98US-0103749.	
PR	09-OCT-1998; 98US-0103794.	
PR	09-OCT-1998; 98US-0103796.	
PR	25-FEB-1999; 99US-0121528.	
XX		
PA	(CHIR) CHIRON CORP.	
PA	(GENO-) INST GENOMIC RES.	
XX		
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;	
PI	Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;	
XX	Tettelin H, Venter JC;	
DR	WPI; 2000-062150/05.	
DR	N-PSDB; AAZ54298.	
XX		
PT	Novel Neisserial polypeptides predicted to be useful antigens for	
PT	vaccines and diagnostics -	
XX		
PS	Claim 2; Page 1211-1212; 1453pp; English.	
XX		
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941	
CC	represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides	
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent	
CC	PCR primers used in the exemplification of the present invention. The	
CC	polypeptides, the polynucleotides, antibodies and compositions of	
CC	the invention can be used as vaccines, as diagnostic reagents, and as	
CC	immunogenic compositions. The polypeptides can be used in the	
CC	manufacture of medicaments for treating or preventing infection due to	
CC	Neisserial bacteria (e.g. meningitis and septicemia), to detect the	
CC	presence of Neisseria bacteria, or to raise antibodies. They may also	
CC	be used to screen for agonists or antagonists, which may themselves	
CC	have use as antibacterial agents. The polynucleotides of the invention	
CC	may also be used in gene therapy protocols.	
XX		
SQ	Sequence 188 AA;	
Query Match	5.6%; Score 213; DB 21; Length 188;	
Best Local Similarity	27.7%; Pred. No. 5.8e-11;	
Matches	51; Conservative 42; Mismatches 75; Indels 16; Gaps 3;	
Qy	8 FRINMTAATVLAALSSS-----VFRAQTADLETVHIKQSYNAIVTEKNGDYSSFAV 60	
Db	6 FSLKILTVMLLSAYGGSPADGWFPVSDGNTVSLDTNVNRGSHALSG-KTEKTRSYTIDRM 64	
Qy	61 TVGKIPASLREIPQSVSIITNQVKDRNVDFDOLARKTPGLRVLSNDDGRSSVYARGY 120	
Db	65 STATGMRIAGKDTFQSVSVITRSLRDKKAVHTLEAMKNTTGVNVNRDSSGLOTRFLSRGP 124	
Qy	121 EYSEYNIDGLPAQMOSING-----TLPNLFAFDRVEVMRGPGLDFSSGEMGGINL 172	
Db	125 YIDQIGEDGITVNVAGRSGYTAKIDVSPSTDLAYVDHIEVVRGATGLTQSNSEPGCTVNL 184	
Qy	173 VRKR 176	
Db	185 IRKR 188	

```

Qy      119 GYEXSEYNIDGLPAWOMOSING-----TLPLNLPAPDRFVFMGEGPLGDFPDSCEMGIV 170
       |||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      123 GFYIDQLDGEDMTNVAVAGRSGETAKIDVSPETSDLAAYVDHIEVVAGATGLTQSNSEPGTV 182
       |||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      171 NLVRK 175
       |||:|||
Db      183 NLRK 187

RESULT 39
ID      AAU29336
AAU29336 standard; Protein: 774 AA.
XX      AAU29336;
AC      AAU29336;
XX      18-DEC-2001 (first entry)
DT      18-DEC-2001 (first entry)
XX      Novel mar regulated protein (NIMR) #8.
DE      mar regulated polypeptide; NIMR; microbial infection; antibacterial.
KW      Escherichia coli.
XX      OS
XX      WO200170776-A2.
XX      27-SEP-2001.
PD      27-SEP-2001.
PF      08-MAR-2001; 2001WO-US07478.
XX      10-MAR-2000; 2000US-188362P.
PR      10-MAR-2000; 2000US-188362P.
XX      (TUFTS ) TUFTS COLLEGE.
PA      Levy SB, Barbosa TM, Alekshun MN;
XX      WPI: 2001-602769/68.
DR      N-PDSB: AAS46238.
XX      Identifying compounds that modulate a newly identified mar regulated
PT      polypeptide activity, useful as antimicrobial compounds, involves
PT      contacting the polypeptide with a test compound -
XX      Disclosure; Page 184-187; 526pp; English.
PS      The invention relates to a method of identifying compounds that modulate
XX      a newly identified mar regulated (NIMR) polypeptide activity. The method
CC      comprises contacting an NIMR polypeptide with a test compound under
CC      interaction conditions, determining the ability of the compound to
CC      modulate the activity or expression of the polypeptide, and selecting the
CC      modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC      of microbial infections, and in screening for modulators of NIMR
CC      expression and activity. These modulators can be used to reduce the
CC      infectivity of a microbe on a surface, and the virulence of a microbe in
CC      a subject suffering from an infection. AAU29329-AAU29379 represent
CC      Escherichia coli NIMR amino acid sequences of the invention.
XX      Sequence 774 AA;
SQ

Query Match          5.4%; Score 205; DB 22; Length 774;
Best Local Similarity 20.1%; Pred. No. 3e-09;
Matches 170; Conservative 108; Mismatches 316; Indels 250; Gaps 39;

Qy      18 LALSSSYFAAO-----PADLETVAIKQRSTNAVTENKGX- 55
       |||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      24 LAGLSFAFAAQNVPAGSLDKALNQVAHSGFTLVSADSTRGQS--NG---HGVD 78
       |||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      56 -----SSFAVTV-----GTKIPASIREIPOSVS 78
       |||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      79 VESGIQLLDSSGLQVKRLGNNSWTLEPAPA PKEDALTIVGDWDIGDARENDFEHHAGARD 138
       |||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      79 IITMQGVKRVNDVTFDQIARRTPGILRVLSNDGGRSSVYARGYESEYN-----ID 128
       |||:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

```
Db 139 VIRREDFAKTATMREVLNRIPIGVSAENNGTSGHDLAMNFGIRGLNPLRASRSTVLMD 198
Qy 129 GLPA-----QMOSINGTLPLNLPFDRVEVMRPGSLFDSSGEMGIVNLVRKRPTKA 180
Db 199 GIPVPFAPYGPQSLAPVSLGNM---DAIDVVRGGAVRYGPQSGVGVNFV-----TRA 251
Qy 181 FQGHAAAGFGTHQYKAAADYSGSLNSDGSVRGRVMAQTVGASPPRAEKNRHEH----- 235
Db 252 IP-----QDFGIEAGVEGQLSPSTSS-----ONNPKEHNLMLV 283
Qy 236 -----FYA---AADWDIN-----PDTVLGAGVLYQOHR-----LAPYNG---L 267
Db 284 GGTADNGFGTALLYSGTRGSDWREHSATRIDDLMLKSKYAPDEVHTFNSLLQYIDGADM 343
Qy 268 PA-----DANNKLPSLPQHVFGADNKNFKQNS--HDVFADLKHFGNGGYGKVGMYR 318
Db 344 FGLSLRADYDADRQWQSTRPYDRF-----NGRRKLASLGQYQFQDSQHKFNIQGFYQTTLR- 398
Qy 319 SDRDADSNVAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPERL 378
Db 399 -----SGLEQGRKITL-----SPRNYWVRGIEPRYSQIEMI 430
Qy 379 GNTANFVIGADYNFRSTNEQRTTLVARGGLALNEFRSIPQVDLIANARKGVRGYSHT 438
Db 431 GPSAHEVGVGYRYLN-ESTHEMYRYTATSSGQLPSG---SSP---YDRTRSGTEAHAW- 482
Qy 439 VATENLDFGYYGKSTFHPADGLSLIGGRGLGHYKIESGEGK---TLHKASKTKFTGYA 494
Db 483 ---YLDKIDIGNWTITP-----GMRPEH--IESVQNNAITGTHEEVSYNAPLEAL 528
Qy 495 GAVYDLNNSLYL-----SLSOLYTPQTNLDADGKLLKPRQNGQEVGKSYMDRLNA 550
Db 529 NVLYHLTDSWNLNANTEGSFGTVQYSQIGKAVQSGNVPEPEKARTWELGTR--YDDGALTA 586
Qy 551 RVSFYRMKDKNAAAFLNPNKKTTRYAALGKRVMEGVETEIS---GAVTP---KWOIHAGY 604
Db 587 ENGLFLINFN---QYDSNQTDVTVARGKTRHTGLETQARVYDLTGLTFLDNLVSIYASY 643
Qy 605 SYLHSQIKTASNSRDDGIFLLPKHSANL-----WTTYQVTPPELTIGGGVNMAMSGIT 656
Db 644 AYVNAIREKGDYGN-LVPFSPKHKGLGVDPKGNW-TFNLNSDFOSSQPADNANTVK 701
Qy 657 SSAGHAG---GYATFDMAAARFTPK---LKQINADNIFNRHYVARVGGANTFNI-PG 709
Db 702 ESADGSTGRIPGFMWGARVAYDFGQWADNLAFGVKNIFPDQYFIRSYDDNNKGIYAG 761
Qy 710 SERT 713
Db 762 QPRT 765

RESULT 40
ABG17860
ID ABG17860 standard; Protein; 454 AA.
XX
AC ABG17860;
AC
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17851.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
```

```
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS82047.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 48219; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 454 AA;
Query Match 5.2%; Score 196; DB 22; Length 454;
Best Local Similarity 24.7%; Pred. No. 8.4e-09;
Matches 74; Conservative 52; Mismatches 127; Indels 46; Gaps 11;
Qy 10 INMTAATVLAALSSVFAAQADLETVHIKQORSYNAIVTEKNGDYSSFAVTGKIPAS 69
Db 55 VAVVAVATAVSGMSVYAQAAPVEPKEDTITVTAAPAPQESAMGPAATIAARQSATGKTDTTP 114
Qy 70 LREIPQSVSIITNQVKDRNVDTFDQLARKTPGLRVLNSDDGRSSVY-----ARGY 120
Db 115 IOKVPQSVISVVTABEMALHQPKSVKELSYTPGVSV--GTRGASNTYDHLIRGPAESGQ 172
Qy 121 EYSEYNIDGLPAQMOSINGTLPLNLPFDRVEVMRPGSLFDSSGEMGIVNLVRKRPT-- 178
Db 173 SONNY-LNGLKLGPNFYNDVIDPYMLERAEMRGPVSVLYGKSNPGCLLNWVSKRPTTE 231
Qy 179 --KAFQGHAAAGFGTHQYKAAEADVSGSLNSDG--SVGRVMAQTVGASP----- 224
Db 232 PLKEVQPKA----GTDSLFTQGFDFDSLSLDDGVVSYRLTGLARSANAQKGSSEQRYAI 287
Qy 225 -----RPAEKNRHEHETFYAADWDINPDTVLGAGVLYQOHRHAPY-NG--L 274
Db 288 APAFTWRPDDKTN-----PTFLUSYFQNEPDTGY-YGWLPEKGTVEPLPKGLRPTDFNDR 341

RESULT 41
AAB40909
ID AAB40909 standard; Protein; 134 AA.
XX
AC AAB40909;
AC
DT 08-FEB-2001 (first entry)
XX
```


CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 134 AA;
Query Match 5.2%; Score 195; DB 23; Length 134;
Best Local Similarity 39.0%; Pred. No. 1.5e-09;
Matches 46; Conservative 20; Mismatches 38; Indels 14; Gaps 3;
QY 457 PADGLSLIGGGRLGHYKIESGCKT-LHKASKTK-----FTGVAGAYVDLNDNNS 505
DB 5 PDDLVLILGTRVSNF---SGTDNTDFDPTKADNRLTYRQGVVTPYAGIYVDLNDIWS 61
QY 506 LYLISLSQLYTPQTNLDADKLLKPRQGNQFVGVKSYMDRLNARVSFYRMKDKNAA 563
DB 62 VYTSYTKIYKPNQSKADRLKLLDPIEGDTYEAGLKAAPFDGLRLNASFAAFRIEQDNVA 119

RESULT 43
ABG29752
ID ABG29752 standard; Protein; 340 AA.

XX AC ABG29752;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #29743.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; RAS93939.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 20; SEQ ID No 60111; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 340 AA;
Query Match 5.2%; Score 195; DB 22; Length 340;
Best Local Similarity 27.1%; Pred. No. 6.6e-09;
Matches 48; Conservative 26; Mismatches 69; Indels 34; Gaps 4;
QY 528 KPRQGNQFVGVKSYMDRLNARVSFYRMKDKNAAAPLNPNKKTRYAALGKRVMEGVE 587
DB 56 KPQKANTSEITGKQVLDKRLLLTAALFRDTIENEV-----EQNDGTYSGYKGRKVEGYE 111
QY 588 TEISGAVTPKQIHHAGYSYLSHSQIKTASNSRDDGIFLL--MPKHSANLWTTTYQVTPELTI 645
DB 112 ISVAGNITPAQVIGGYTQOKATIKNGKDVADQSSSLPYTPEHAFTLMSQYQATDDISV 171
QY 646 GGVNAMSGITSSAGMHAGGYATFDMAAAYRFTPKLKLQINADIFNRHYVARVGA 702
DB 172 GAGARYI-----GSMHKG-----SDGANNETYGFHVDGA 200

RESULT 44
AAY34494
ID AAY34494 standard; Protein; 833 AA.

XX AC AAY34494;
XX DT 25-AUG-1999 (first entry)
XX DE Porphyromonas gingivalis protein PG35.
XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX KW vaccine; antigenic.
XX OS Porphyromonas gingivalis.
XX PN WO9929870-A1.
XX PD 17-JUN-1999.
XX PF 10-DEC-1998; 98WO-AU01023.
XX PR 04-AUG-1998; 98AU-0005028.
XX PR 10-DEC-1997; 97AU-0000839.
XX PR 31-DEC-1997; 97AU-0001182.
XX PR 30-JAN-1998; 98AU-0001546.
XX PR 10-MAR-1998; 98AU-0002264.
XX PR 09-APR-1998; 98AU-0002911.
XX PR 23-APR-1998; 98AU-0003128.
XX PR 05-MAY-1998; 98AU-0003338.
XX PR 22-MAY-1998; 98AU-0003654.
XX PR 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.
XX PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
XX PI Ross BC, Rothel LJ, Webb EA;
XX DR WPI; 1999-385613/32.
XX DR N-PSDB; AAX91712.

XX PT Antigenic Porphyromonas gingivalis peptides for preventing
XX PT gingivitis
XX PS Claim 1; Page 474-475; 588pp; English.
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial

CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

XX Sequence 833 AA;

Query Match 5.1%; Score 193.5; DB 20; Length 833;
 Best Local Similarity 20.4%; Pred. No. 3.7e-08;
 Matches 142; Conservative 82; Mismatches 266; Indels 207; Gaps 32;

QY 34 EYVHIK--GQSYNAIVTEKNGDYSSFAVTGKIPASLREIPQSVSIITNOQVQRNVD 91
 DB 88 EKVIHKGSRHVDLYLTELISLDGVVVS--ANRNETFRQAPSLVTVLSPELFLKTNST 146
 QY 92 TFDQLARKTPGLRVLSN--DDGRSSVYARGYE--YSE-----YNIDGLPA 132
 DB 147 NISQGLKFPQGLRVEDNCCQGFNOVRINGLEGAYSQIILDSHPITSSLAGVGLQMPA 206
 QY 133 QMOSINGTLPNLFAPRVEVMRGPGLFDSSGEMGIVNLVRKRPVK--AFQGHAAAGFG 190
 DB 207 NM-----IERVEYIRGGGSALEFGSNVGVGINVITEPLRNSAEISHSTMTFD 254
 QY 191 THKOYRAEADVSGSLNSDSVGRVMAQTVGASPRPAEKNNRHETTYAADMWDINDPTVL 250
 DB 255 HAKGM-----GSFQNTQOFNGSMLTE-----DRKA 279
 QY 251 GAGYLKQQRHLAPYNLPPADANNKLSLPQHVFGADMKFKKNSHDVPADLKHFGNGG 310
 DB 280 GV-MVFGQHNYRPGQDIDGNFTLPLNRSL-----GFRSYKKTGL 321
 QY 311 YGKVGKRYSDRDADSNYAFAAGSKLGMKTPAGRPGCNTADKACAVGLTEIKQKALAFDA 370
 DB 322 YSKATLEIYHSM---QYRRGGDRLD-----NPPFAQIAEYLQHYINGGSKFEDQ 368
 QY 371 SYSRP--FRIGTANE---FVIGADY--NFRSTNEQRTTL--YARGIALNEFRS 418
 DB 369 GREGGKDFSLYASQDVGRSRYGGDYEENLNGAVQSGSTESDEYNDAPFALTSYGT 428
 QY 419 IPQVDLIANARKGVRGYSHTVATENDEFG---IYK-----STPHPA 458
 DB 429 YKGFEDL---QGGMYRHTFG--ENWDFTGLEYIGQLDRSGYRPSKIDONTSTFSQY 482
 QY 459 DGLSLIGGRGLGHYKIESGEGKTIHAKSKTFTGYAGAYVD---LNDNNSLYLSQLYT 515
 DB 483 DQI-----EYKTB-----KLSALIGARIDYVLANDGKRYIDPLEPFS 520
 QY 516 PQTNLDADGKLLKPRQGNQFVEYK-----GSYMDRLNARVSPYRKDKNAAP---L 566
 DB 521 PRANV-----RYPNKNLSPRLSYSEGFRAQYFDEDLHVELA-----GGTPISRVL 567
 QY 567 NPNNKTRVLAALGK-----RVN--EGVETFEISAVTPKMOIHAGYISLSQIK 612
 DB 568 SPMLKEERSISIASPDYHRADEWFNIMGELAFSTFISNQFKRSDKVEVTTSDGKEWIR 627
 QY 613 TASNRRDDGIFILMPKXANL--WTTYQVTPPELTIG 647
 DB 628 TIYNKDG---VSKYGVNLEGRIVAKNSFDLQGG 660

RESULT 45

AAV34368
 ID AAV34368 standard; Protein; 876 AA.

XX AAV34368;

XX 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG35.

XX Porphyromonas gingivalis; PG, periodontal disease; gingivitis;

KW vaccine; antigenic.

XX Porphyromonas gingivalis.

OS WO9292670-A1.

XX 17-JUN-1999.

XX 10-DEC-1998; 98WO-AU01023.

XX 04-AUG-1998; 98AU-0005028.

XX 10-DEC-1997; 97AU-0000839.

XX 31-DEC-1997; 97AU-0001182.

XX 30-JAN-1998; 98AU-0001546.

XX 10-MAR-1998; 98AU-0002264.

XX 09-APR-1998; 98AU-0002911.

XX 23-APR-1998; 98AU-0003128.

XX 05-MAY-1998; 98AU-0003338.

XX 22-MAY-1998; 98AU-0003654.

XX 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;

XX Ross BC, Rotnel LJ, Webb EA;

XX WPI; 1999-385613/32.

XX N-PSDB; AAX91586.

XX Antigenic Porphyromonas gingivalis peptides for preventing

XX gingivitis

XX Claim 1; Page 330-331; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34338 to

XX AAV34383. AAX91802 to AAX91989 represent PCR primers used in the

XX isolation of the PG polypeptides. The PG polypeptides have antibacterial

XX activity with a vaccine mechanism of action. The PG polypeptides can be

XX used as vaccines especially against Porphyromonas gingivalis. Probes can

XX be used to detect Porphyromonas gingivalis in standard hybridisation

XX assays. Porphyromonas gingivalis is involved in periodontal disease

XX especially gingivitis.

XX Sequence 876 AA;

Query Match 5.1%; Score 193.5; DB 20; Length 876;
 Best Local Similarity 20.4%; Pred. No. 4e-08;
 Matches 142; Conservative 82; Mismatches 266; Indels 207; Gaps 32;

QY 34 EYVHIK--GQSYNAIVTEKNGDYSSFAVTGKIPASLREIPQSVSIITNOQVQRNVD 91
 DB 131 EKVIHKGSRHVDLYLTELISLDGVVVS--ANRNETFRQAPSLVTVLSPELFLKTNST 189
 QY 92 TFDQLARKTPGLRVLSN--DDGRSSVYARGYE--YSE-----YNIDGLPA 132
 DB 190 NISQGLKFPQGLRVEDNCCQGFNOVRINGLEGAYSQIILDSHPITSSLAGVGLQMPA 249
 QY 133 QMOSINGTLPNLFAPRVEVMRGPGLFDSSGEMGIVNLVRKRPVK--AFQGHAAAGFG 190
 DB 250 NM-----IERVEYIRGGGSALEFGSNVGVGINVITEPLRNSAEISHSTMTFD 297
 QY 191 THKOYRAEADVSGSLNSDSVGRVMAQTVGASPRPAEKNNRHETTYAADMWDINDPTVL 250
 DB 298 HAKGM-----GSFQNTQOFNGSMLTE-----DRKA 322
 QY 251 GAGYLKQQRHLAPYNLPPADANNKLSLPQHVFGADMKFKKNSHDVPADLKHFGNGG 310
 DB 323 GV-MVFGQHNYRPGQDIDGNFTLPLNRSL-----GFRSYKKTGL 364
 QY 311 YGKVGKRYSDRDADSNYAFAAGSKLGMKTPAGRPGCNTADKACAVGLTEIKQKALAFDA 370
 DB 365 YSKATLEIYHSM---QYRRGGDRLD-----NPPFAQIAEYLQHYINGGSKFEDQ 411

Search completed: December 18, 2002, 06:40:21
Job time : 53.9653 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:40:49 ; Search time 16.862 Seconds
(without alignments)
1259.836 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776
Sequence: 1 MGQMSVPRINMTAATVLA.....NTFNIPGSERTWTNIRYSF 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
 - 2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
 - 3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
 - 4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
 - 5: /cgn2_6/prodata/1/1aa/6CCTS_COMB.pep:*
 - 6: /cgn2_6/prodata/1/1aa/backfillset1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296.5	7.9	696	1	US-08-765-081-5 Sequence 5, Appl
2	296.5	7.9	696	3	US-09-098-082-5 Sequence 5, Appl
3	296.5	7.9	696	5	PCT-US95-06994-5 Sequence 5, Appl
4	286	7.6	703	5	PCT-US95-06994-8 Sequence 8, Appl
5	273.5	7.2	725	4	US-09-668-113A-8 Sequence 6, Appl
6	264.5	7.0	718	5	PCT-US95-06994-6 Sequence 6, Appl
7	260	6.9	663	1	US-08-765-081-7 Sequence 7, Appl
8	260	6.9	663	3	US-09-098-082-7 Sequence 7, Appl
9	232.5	6.2	703	5	PCT-US95-06994-9 Sequence 9, Appl
10	212	5.6	652	3	US-08-765-081-6 Sequence 6, Appl
11	212	5.6	652	3	US-09-098-082-6 Sequence 6, Appl
12	186.5	4.9	791	3	US-08-537-361E-4 Sequence 4, Appl
13	186.5	4.9	791	4	US-08-817-707-4 Sequence 4, Appl
14	185	4.9	792	1	US-08-326-670A-2 Sequence 2, Appl
15	183	4.8	792	3	US-08-990-470A-2 Sequence 2, Appl
16	183	4.8	792	4	US-08-817-707-2 Sequence 2, Appl
17	181	4.8	718	5	PCT-US95-06994-7 Sequence 7, Appl
18	180.5	4.8	718	5	PCT-US95-06994-7 Sequence 7, Appl
19	180	4.8	792	3	US-08-537-361E-2 Sequence 2, Appl
20	179	4.7	792	3	US-08-537-361E-6 Sequence 6, Appl
21	173	4.6	713	4	US-08-613-009A-18 Sequence 18, Appl
22	173	4.6	713	4	US-08-613-009A-18 Sequence 18, Appl
23	173	4.6	915	4	US-08-778-570B-24 Sequence 24, Appl
24	173	4.6	915	4	US-09-059-584-24 Sequence 24, Appl
25	172	4.6	915	1	US-08-487-890A-96 Sequence 96, Appl
26	172	4.6	915	2	US-08-363-124A-2 Sequence 2, Appl
27	172	4.6	915	2	US-08-478-435-96 Sequence 96, Appl

28	172	4.6	915	2	US-08-337-483-96 Sequence 96, Appl
29	172	4.6	915	2	US-08-478-373-96 Sequence 96, Appl
30	172	4.6	915	3	US-08-474-671-96 Sequence 96, Appl
31	172	4.6	915	3	US-08-483-577A-96 Sequence 96, Appl
32	172	4.6	915	4	US-08-897-438-96 Sequence 96, Appl
33	172	4.6	915	4	US-08-637-654-96 Sequence 96, Appl
34	172	4.6	915	4	US-08-649-518-96 Sequence 96, Appl
35	171	4.5	909	2	US-08-363-124A-4 Sequence 16, Appl
36	169.5	4.5	790	4	US-08-817-707-8 Sequence 16, Appl
37	169.5	4.5	908	3	US-08-613-009A-16 Sequence 22, Appl
38	169.5	4.5	908	4	US-08-778-570B-22 Sequence 22, Appl
39	169.5	4.5	908	4	US-09-059-584-22 Sequence 94, Appl
40	167.5	4.4	908	1	US-08-487-890A-94 Sequence 94, Appl
41	167.5	4.4	908	2	US-08-478-435-94 Sequence 94, Appl
42	167.5	4.4	908	2	US-08-337-483-94 Sequence 94, Appl
43	167.5	4.4	908	2	US-08-478-373-94 Sequence 94, Appl
44	167.5	4.4	908	3	US-08-474-671-94 Sequence 94, Appl
45	167.5	4.4	908	3	US-08-483-577A-94 Sequence 94, Appl

ALIGNMENTS

RESULT 1
US-08-765-081-5
; Sequence 5, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.R., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSER: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-765-081-5
Query Match 7.9%; Score 296.5; DB 1; Length 696;
Best Local Similarity 21.3%; Pred. No. 1.1e-18;
Matches 176; Conservative 103; Mismatches 304; Indels 243; Gaps 39;
QY 9 RIMNTAATVLAASSVFAQQTADLETVAHKGRSTNAIVTEKNGYSSFAVVGKIPR 68
DB 2 RITTLASVPIPLGFGS-ASSIAADVMIIVASGY-----EK----- 37

```

QY 69 SUREIPQSVSIITNOQVKDRNVDTFDQARKTPGLRVLS--NDDGRSSVYARGY--EYSE 124
DB 38 KLTNAASVSIVSQBELSQSHYDLAEALRSVEGVDSGTGKTGLEISIRGMPASVTL 97
QY 125 YNIDGL-----PAQMOSIN-GTLPNLFADRVEMRGPSGLFDSSGEMGIVNLVR 174
DB 98 ILIDGVRQGGSDVTPNGFSAMNTGMPPLAAIERIEVIRGPMSTLYGSDAMGVVNIIT 157
QY 175 KRPTKAFQGHAAAG-----FGTHKQYK-----AEADVSGSLNSDGSVRGR-----V 215
DB 158 RKNADKWLSSVNAGLNQSNNKWNSSQNFWSGGLVDDSVLSQVRGSTQQRQSSVTS 217
QY 216 MAQTVGAS--PRPAEKNNRHETFYAAADW-----DI-----NPDTVLGA---G 253
DB 218 LSDTAGTRIPYPTESQNYN--LGAFLDKWAKSEQDVLHFDMDTTRQYDNRDQGLSLTGG 275
QY 254 YLYQORHLAPYNGLPADANNKLPSPQHVFGADWNKFKMNSHDVPADLKHVFGNGGYGK 313
DB 276 YDRTLRY-----ERNKISAGYDHTFTFTGWTXSY-LNNWNET-----ENKREL 316
QY 314 VGMRYSDRDADSNVAFAGSKLGMKTPAGRPCNCNTADDKACAVGLGTEIKOKALAFDASYS 373
DB 317 V-----RSVLKRDWKGL---AGQP-----RELKESNLILNSLLL 347
QY 374 RPPRLGNTANEFVIGADYNRFRSTNEOGRITLYARGGLALNEFRSIPQVDLIANARKGVR 433
DB 348 TPL-----GESHLVTVGG-----EFQSSSMKDGCVVLASTG-- 377
QY 434 GYSHTVATENLDEFGIYKSTPHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKFTGY 493
DB 378 -----ETFRQKSWSVFAEDEWHLTDALALTAGSYEHHEQFGH-----FSRP 420
QY 494 AGAVYDLNDNNSL-----YLS--LSQLYTPQTNLDADGKL-----LKPRQNGQFEV 537
DB 421 AYLWMDVADAWTLKGGVTTGYKAPRMGQLHKGISGVSGQKTNLLGNPDLPKPEESVYEA 480
QY 538 GYKGSYMD--RLNARVSFYRMKDKNAAPLNPNKTKRYAALGKRVMEGVETEISGAVT 595
DB 481 GV---YDNPAGLANVTGFMDFSNKIVSYINDNTNSVYNSGKARLHGV--EPAGTL- 534
QY 596 PKWQ-----IHAGSYLHSGIKTASNRDGDIFLLPKHSANLWTTVQYTPELTIGGV-- 649
DB 535 PLWSEDTLSLNTWTRSEQRDGNK--GAPLSYTPHEHVMNAKLWQITEEVASWLGARY 592
QY 650 -----NAMSGITSSA-----GMHAGGYATFDMAAAYRFTPKLKLQINADNIFNRH 694
DB 593 RGKTRFRFTQNYSSLASVQKVDYDEKGEYLKAWTVVDAGLSWKMTDALTILNAAVNLLNKD 652
QY 695 Y-----YARVGGANT-FNIPGSBRTWTANLRYSF 722
DB 653 YSDVSLYSAGKSTLYAGDYFQTGSSTGYVIP--ERNYMWLSLNYQF 696

```

RESULT 2

US-09-098-082-5

Sequence 5, Application US/09098082

Patent No. 6040421

GENERAL INFORMATION:

APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.

TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESS: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

COMPUTER: IBM PC compatible/Pentium II

OPERATING SYSTEM: MS-Windows 95

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/098,082

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/765,081

FILING DATE: March 26, 1997

APPLICATION NUMBER: PCT/US95/06994

FILING DATE: June 7, 1995

APPLICATION NUMBER: US 08/265,714

FILING DATE: June 24, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: CHOR-1-12402

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)

TELEFAX: 1-206-224-0779

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 696 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-098-082-5

Query Match

Best Local Similarity 21.3%; Pred. No. 1.1e-18;

Matches 176; Conservative 103; Mismatches 304; Indels 243; Gaps 39;

QY 9 RINTATVLAALSSVFAAQTADLETVHKGQSYNAIVTEKNGDYSSPAVTVGTKIPA 68

DB 2 RITTLASVVPCLGFS--ASSIAAAEDVMIVSASGY-----EK----- 37

QY 69 SLREIPQSVSIITNOQVKDRNVDTFDQARKTPGLRVLS--NDDGRSSVYARGY--EYSE 124

DB 38 KLTNAASVSIVSQBELSQSHYDLAEALRSVEGVDSGTGKTGLEISIRGMPASVTL 97

QY 125 YNIDGL-----PAQMOSIN-GTLPNLFADRVEMRGPSGLFDSSGEMGIVNLVR 174

DB 98 ILIDGVRQGGSDVTPNGFSAMNTGMPPLAAIERIEVIRGPMSTLYGSDAMGVVNIIT 157

QY 175 KRPTKAFQGHAAAG-----FGTHKQYK-----AEADVSGSLNSDGSVRGR-----V 215

DB 158 RKNADKWLSSVNAGLNQSNNKWNSSQNFWSGGLVDDSVLSQVRGSTQQRQSSVTS 217

QY 216 MAQTVGAS--PRPAEKNNRHETFYAAADW-----DI-----NPDTVLGA---G 253

DB 218 LSDTAGTRIPYPTESQNYN--LGAFLDKWAKSEQDVLHFDMDTTRQYDNRDQGLSLTGG 275

QY 254 YLYQORHLAPYNGLPADANNKLPSPQHVFGADWNKFKMNSHDVPADLKHVFGNGGYGK 313

DB 276 YDRTLRY-----ERNKISAGYDHTFTFTGWTXSY-LNNWNET-----ENKREL 316

QY 314 VGMRYSDRDADSNVAFAGSKLGMKTPAGRPCNCNTADDKACAVGLGTEIKOKALAFDASYS 373

DB 317 V-----RSVLKRDWKGL---AGQP-----RELKESNLILNSLLL 347

QY 374 RPPRLGNTANEFVIGADYNRFRSTNEOGRITLYARGGLALNEFRSIPQVDLIANARKGVR 433

DB 348 TPL-----GESHLVTVGG-----EFQSSSMKDGCVVLASTG-- 377

QY 434 GYSHTVATENLDEFGIYKSTPHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKFTGY 493

DB 378 -----ETFRQKSWSVFAEDEWHLTDALALTAGSYEHHEQFGH-----FSRP 420

QY 494 AGAVYDLNDNNSL-----YLS--LSQLYTPQTNLDADGKL-----LKPRQNGQFEV 537

DB 421 AYLWMDVADAWTLKGGVTTGYKAPRMGQLHKGISGVSGQKTNLLGNPDLPKPEESVYEA 480

QY 538 GYKGSYMD--RLNARVSFYRMKDKNAAPLNPNKTKRYAALGKRVMEGVETEISGAVT 595

DB 481 GV---YDNPAGLANVTGFMDFSNKIVSYINDNTNSVYNSGKARLHGV--EPAGTL- 534

QY 596 PKWQ-----IHAGSYLHQSQIKTASNSRBDGIFLLMPKHSANLMTTYQVPELTIGGCV-- 649
 DB 535 PLWSEBVTLSLNTWTWRSRQDGDNK--GAPLSYTPHEHWNNAKLNNQITEBVAWSLGGARY 592
 QY 650 -----NMSGITSSA-----GMHAGGYATFPDAMAAYRFTPKLQIANDNIFNRH 694
 DB 593 RGTPTFTQNYSSLSAVQKKVYDEKEBYLKAMTVVDAGLSWKMTDALTLNAAVNNLNLKXO 652
 QY 695 Y-----YARVGANT-FNIPGSERTWTANLRYSF 722
 DB 653 YSDVSLYSAGKSTLYAGDYFQTGSSSTTGYVIF--ERNYMSLNTYQF 696

RESULT 3
 PCT-US95-06994-5
 Sequence 5, Application PC/TUS9506994
 GENERAL INFORMATION:
 APPLICANT: Children's Hospital & Medical Center
 APPLICANT: University of Washington
 APPLICANT: Washington State University Research Foundation
 APPLICANT: TARR, Phillip I
 APPLICANT: BILGE, SIMA S
 APPLICANT: BESSER, THOMAS E
 APPLICANT: VARY JR, JAMES C
 TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 STREET: SUITE 2800, 1420 FIFTH AVENUE
 CITY: SEATTLE
 STATE: WASHINGTON
 COUNTRY: USA
 ZIP: WA 98101
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/06994
 FILING DATE: 07-JUN-95
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/265,714
 FILING DATE: 24-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: BRODERICK, THOMAS F
 REGISTRATION NUMBER: 31,332
 REFERENCE/DOCKET NUMBER: CHOR-18591
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682 8100
 TELEFAX: (206) 224 0779
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 696 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-06994-5

Query Match 7.9%; Score 296.5; DB 5; Length 696;
 Best Local Similarity 21.3%; Pred. No. 1.1e-18;
 Matches 176; Conservative 103; Mismatches 304; Indels 243; Gaps 39;
 QY 9 RINMTATATLALSSVFPAAQADLTETVHIKQGRSYNAVTEKNGYSSFAVTGKIPA 68
 DB 2 RITTLASVVIPICGFS--ASSISAAEDVMIVSAGY-----EK----- 37
 QY 69 SLREIQSVSIITNOQVKORNVDTFPDLARKTPGLVLS--NDGSSSVYARGV--EYSE 124
 DB 38 KLTNMAASVVISQEBLQSSQYHDLAALBSVGEVVESTGTGGLGLEISIKMPASYYTL 97

QY 125 YNIDGL-----PAQMSIN-GTLFNLPAFDRVYVNRGSEGLFDSSGEMGIVNLVR 174
 DB 98 ILIDGVROGGSDDVTENGSANMTGMPPLAIERIEVIRGEPSTLYGSDMGVNIIT 157
 QY 175 KRPTKAFQGHAAAG-----FGTHQYK-----AEADVSGSLNDGSRGR-----Y 215
 DB 158 RKNADKMLSSVNNAGLNLQSNKKGNSSQFNFWSGGLVDDSVSLQVRGSGTQOQSGSVTS 217
 QY 216 MAQTVGAS--PRPAEKNNRHETFFYAADM-----DI-----NPDYTLGA---G 253
 DB 218 LSPDTAGTRIPYTESQNYN--LGARLDWKASBODVLMFMDTTRQRYDNRDGLGSLTGG 275
 QY 254 YLYQGHLPAPYNGLPADANNKLPSLPQHVFGADWNKFKRNHSHVDPADLKHYFGNGGYGK 313
 DB 276 YDRLTRY-----ERNKISAGYDHTFTFGTWKSY-LNNMET-----ENKGREL 316
 QY 314 YGKRYSDBRADSNYPAGSKLGMKTPAGRPQCGTTADDKACAVLGEIRIQKALAPDASIS 373
 DB 317 V-----RSYKRDWKGL--AGQP-----RELKESNLILNSLL 347
 QY 374 RPPRLGNTANEPVIGADYVNRFRSTNEQGRITLYARGGLALNEFRSIPQYDLINARKGYR 433
 DB 348 TPL-----GESHLVTGG-----EFQSSMDQGVLAATG-- 377
 QY 434 GYSHVATENLDEFGIYKSTFHPADGLSIGGRIGHYKIESGEGKTLHKASKTKFTGY 493
 DB 378 -----ETFRQKSWSVFAEDENHTLTALALTAGSRVHHQFGGH-----FSPR 420
 QY 494 AGAVYLDNNNSL-----YLS--LSQYTPQTNLDADGKL-----LKPRQGNQFEY 537
 DB 421 AYLVMDVADAMTLKGGVTGTGYKAPRMGQLHKGISGVSQGGKTNLLGNPDLKPESSVYEA 480
 QY 538 GYKGSYMD--RLNARVSFPRMKDKNAALPLNPNKKTRYAALGKRVMEGVETEISGAVT 595
 DB 481 GV--YNDNPAGLNNAVTPGMTDFSKYISYSINDNTNSYVANSGRKLRGV--EFGTLL 534
 QY 596 PKWQ-----IHAGSYLHQSQIKTASNSRBDGIFLLMPKHSANLMTTYQVPELTIGGCV-- 649
 DB 535 PLWSEBVTLSLNTWTWRSRQDGDNK--GAPLSYTPHEHWNNAKLNNQITEBVAWSLGGARY 592
 QY 650 -----NMSGITSSA-----GMHAGGYATFPDAMAAYRFTPKLQIANDNIFNRH 694
 DB 593 RGTPTFTQNYSSLSAVQKKVYDEKEBYLKAMTVVDAGLSWKMTDALTLNAAVNNLNLKXO 652
 QY 695 Y-----YARVGANT-FNIPGSERTWTANLRYSF 722
 DB 653 YSDVSLYSAGKSTLYAGDYFQTGSSSTTGYVIF--ERNYMSLNTYQF 696

RESULT 4
 PCT-US95-06994-8
 Sequence 8, Application PC/TUS9506994
 GENERAL INFORMATION:
 APPLICANT: Children's Hospital & Medical Center
 APPLICANT: University of Washington
 APPLICANT: Washington State University Research Foundation
 APPLICANT: TARR, Phillip I
 APPLICANT: BILGE, SIMA S
 APPLICANT: BESSER, THOMAS E
 APPLICANT: VARY JR, JAMES C
 TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 STREET: SUITE 2800, 1420 FIFTH AVENUE
 CITY: SEATTLE
 STATE: WASHINGTON
 COUNTRY: USA
 ZIP: WA 98101
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 703 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
DESCRIPTION: wherein "Xaa" residues represent gaps
DESCRIPTION: introduced to facilitate best alignment with
DESCRIPTION: SEQ ID NO:9.
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
PCT-US95-06994-8

Query Match 7.6%; Score 286; DB 5; Length 703;
Best Local Similarity 20.4%; Pred. No. 1.1e-17;
Matches 171; Conservative 104; Mismatches 304; Indels 258; Gaps 38;
QY 9 RINMTAATVLAAL---SSSVFAAQTADLTVHLKQORSNAIVTEKNGDYSSFAVTVGTRK 65
DB 2 RITTLASVVIPCLGFSASSIXXXXXAAEDVMIVSAGY-----EK----- 42
QY 66 IPASLREIPQSVSIITNQOVKDRNVDTPDLQARKTPGLRVLS--NDDGRSSVYARGY--E 121
DB 43 ---KLTNAAASVSVISQEBLSQQYHDLAELARSVEGVDVESGTGKTGGLSIRGMPAS 99
QY 122 YSEYIDGL-----PAQWQIN-GTLPNLFAPDRVEVMRGPGLPDSGEMGGINV 171
DB 100 YTLILIDVROGGSSDVTENGFSAMTGFMPPLAAIERIEVIRGPMSTLYGSDAMGGVYN 159
QY 172 LVKRPTKAFQGHAAAG-----FGTHQYK-----AEADVSGSLNSDGSVGRG--- 214
DB 160 IITRKNADKWLSSVAGLNLQSNKMGNSQFNWSSGGLVDVSLQVRGSTQORQGS 219
QY 215 --VMAQTVGAS--PRAEKNRHETFYAAADWDINPDVLGAGLYQORHLAPYNGLPADA 271
DB 220 VTSLSDTAGTRIPYPTESQYN--LGARLDWKASEQDVL----- 256
QY 272 NNKLPSLPOHVFVGGADWPKFNKNSHDVFADLKHYPGNGYGVKGMYSRDRDADSNVAFAG 331
DB 257 -----W--FDMDT-----TRQYDNRDQG-----LG 275
QY 332 SKLGMKTPAGRPGCNTADDKACAVGLGTETIKOKALAFADASYSRPRRLGNTANEFVIGADY 391
DB 276 SUTG-----GVDRLRYERNKISAGYDGTGTFPGTWKS-----Y 308
QY 392 NRFRSTNEQGRITLVA-----RGGLALNFRSIPQVDLIANA-RKGVRSYSHVAT----- 441
DB 309 LNWNETENKRELVRSLKRDKWGLA-GQPRELKESNLILNSLLTPLGESHLVTVGGEF 367
QY 442 -----ENLDFYIGKSTPHPADGLSLIGGRGLGHVYKIESGEGKTLH 483
DB 368 QSSSMKDGVVVLASTGTETFRQKXSWVFAEDWEHLTDALATAGSRYEHHEQFGGH----- 422
QY 484 KASKTKFTGYAGAVVDLNDNNSL-----YLS--LSQLYTPQTNLDDADGKL-----L 527

DB 423 -----FSPRAYLVWDVADAWTLKGVTTGYKAPRMGQLHKGISGVGGKTNLLGNPDL 476
QY 528 KPRQNGQFEVYKGSYMD--RLNARVSFYRMKDKN-AAAPLNPNKKTRYAALGRVME 584
DB 477 KPEESVSEAGV--YYDNPAGLNANVTGFTDFSNKIXSVSINDNTNSVYNSGKARLH 533
QY 585 GVETISCAVTPKQKQ-----IHAGYSYLSHLSQIKTASNRDDGIFLLMPKHSANLWTTYQVT 640
DB 534 GV--EFAGTL-PLMSEDTVLSNYTWTRSEQRDGNK--GAPLSYTPHEHVMNAKLNWQIT 588
QY 641 PELTITGGGV-----NAMSGITSSA-----GMHAGGYATFDAAAYRFTPKLKL 683
DB 589 EVASWMLGARYRGKTPRTQNYSSLSAVQKVYDEKGYLKAWTVVDAGLSKWTDALTL 648
QY 684 QINADNIFNRHY-----YARVGANT-FNIPGSERTWTANLRYSF 722
DB 649 NAAVNNLLNKDYSVLSYAGSAGTLYAGDYFQTSSTGYVIP--ERNYWSLNTQF 703
RESULT 5
US-09-668-113A-8
Sequence 8, Application US/09668113A
Patent No. 6410703
GENERAL INFORMATION:
APPLICANT: Russo, Thomas A.
TITLE OF INVENTION: Identification of A Vaccine Candidate from an
EXTRINSIC STRAIN OF E. COLI
FILE REFERENCE: 11520.0214
CURRENT APPLICATION NUMBER: US/09/668,113A
CURRENT FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 10
SEQ ID NO 8
LENGTH: 725
TYPE: PRT
ORGANISM: Escherichia coli
FEATURE:
US-09-668-113A-8
Query Match 7.2%; Score 273.5; DB 4; Length 725;
Best Local Similarity 19.8%; Pred. No. 1.7e-16;
Matches 167; Conservative 127; Mismatches 304; Indels 247; Gaps 40;
QY 5 MSVFRINMTAATVLAALSSSVFAAQTADLTVHLKQORSNAIVTEKNGDYSSFAVTVGT 64
DB 1 MRINKILWSLTVLLVGLNSQVSAKYS-----DNDDELTVVEA 39
QY 65 KIPASLREIPQSVSIITNQOVKDR-NVDTPDLQARKTPGLRVLS-----DGRSSVY 116
DB 40 TAEQVLKQOP-GSVVITSEDIKTPPVNDLSDIIRKMGFVNLGNSASGTRGNRQIDIR 98
QY 117 ARGVEYSYINIDGLPAQMOS-----INGTLPNLFAPDRVEVMRGPGLPDS 162
DB 99 GMPENTLILIDGVPTSRNSVYSWRGERDTRGDTNRWVPP-EQVERIEVIRGPAARVY 157
QY 163 SGEMGIIVNLVRKPTKAFQGHAAAGFTGHQYKAE-----DVSGSLNSDG--SV 211
DB 158 SGAAGGVVNIITKTRPTNDW--HGSLSLYTNQPESEEGATRANFSLSGPLAGDALITRL 215
QY 212 RGRV-----MAQTVGASPRPAEKNRHETFYAAADWDINPDVLG--AGLYLQQRH 260
DB 216 YGNLKTADSDWDINS PVGTGKNAAGHEGVRNKDINGVSWKLNPOOILDFEYVSRQG-- 273
QY 261 LAPVNGLPADANNKLSLPQHVFV--GADWNKFKNNSHDVFADLKHYPGNG-----GYGKV 314
DB 274 ---NIYAGDTQNSSSSAVTELSAKSGKETNRLYQNYGI---TH---NGIWDWGQSRF 322
QY 315 GMRYSRDRDADSNVAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTETIKOKALAFADASYSR 374
DB 323 GYVYEKTN-----NTRMNEGLSGGEGEIRLAGEK-----FTTNRLS 358
QY 375 PFRLGNTAN-----EFVIGADYNRFRSTNEQGRITLVARGGALNFRSIPQVDLI 425
DB 359 SWRTSGELNIPLVNVDQTLTVGAENNRDKLDDPSSTS-----LTVND-RDI----- 404

QY 405 LYARGGLALNERSIPQVDLIANARKVGRYSHTVATENLDFEGLYKGS-----TTHP 457
DB 390 QVAL--FVEDEWRIEPEPLATT--GVR-----MDDHETYGHEWSPRAYLVNA 433
QY 458 ADGLSLIGGRGLGHVKGESGEGKTLHKASKTKFTGYAGAVYDLNNSLYLSLSQLYTPQ 517
DB 434 TDTVTVTKG-----WATAFKAP-----SLLQLSPDW 459
QY 518 TNLADAGKL-----LKPRQGNQPEVG--YKGS-----YMDRLNARVSFY 555
DB 460 TNSCRGACKIVGSPDLKPETSESWEGLIYMGEEGLEGVSSVTVFEND-VKDRIGIS 518
QY 556 RMKDNAAAPLNPNKTKRYAALGRV-----MEGVETEISGAVTPKQIHAQ 603
DB 519 RTSDEVN-AAPGQNFVGFETGANGRRIPVFSYNNYNNKARNQGVETELKIPFNDWKLSIN 577
QY 604 YSLHSQIKTASNRDDGIFLLMPKHA-NLWTTVQVTPELTIGGVNAMSGITSSAGMH 662
DB 578 YTY--NDGRDVNGENKPLSDI--PFLALEDWSFY-----VSGHYTCOKRADSATAKT 627
QY 663 AGGYATFDMAAAYRFTPKLKLQINADNI 690
DB 628 PGYTIWNTGAAWQTKDVKLRAGVLNL 655

RESULT 10

US-08-765-081-6
; Sequence 6, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrgA amino acid sequence
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Vibrio Cholerae
US-08-765-081-6

Query Match 5.6%; Score 212; DB 1; Length 652;

Best Local Similarity 19.6%; Pred No. 7.4e-11;
Matches 147; Conservative 99; Mismatches 299; Indels 204; Gaps 32;
QY 57 SPAVTVGTKIPAS-----LREIPQSVSIITNQOQKDRNDVTF 93
DB 10 SLSVTTLGLMFSASAFQADATKTDVTMVTAAGYAQVIONAPASISVISREDLESRYRDV 69
QY 94 DQARKTPGLRLVLSNDGRS--SVYARGYEVSEYINIDGLPAQMOSI-----NGTLP 142
DB 70 TDALKSPGVTVTGGDTTIDISIRMGSGNYTLILVDGKROTSTRTNPSDGPGEQGWLP 129
QY 143 NLPAFDRVEMVRGSGFLFDDSGEMGGIVNLVRKPTTAFQGHAAAGCTHKKYKAEADVS 202
DB 130 PLOAERIEVIRGPMSTLYGSDAIGGVINIITRKQQQWMSGNVOLSTVVQENRASGDQES 189
QY 203 GSLNSDGSVRGRVMAQTVGASPRPAEKNNRH-----ETFYAADMDINDP--TVLGAG 253
DB 190 ANFFVTGPLSDALSILQVYVGTQTTORDEDEIEHGYGDKLSRLTSKLYQLNPDHQLQLEAG 249
QY 254 YLYQORH-----LAPYNGLPADANNKPLSLPQHVFVGADWNKFKON-----SHDVPADLKH 304
DB 250 VSAQDRENNVGKSAQSGCGRTCSN-----TD-NQYRRNHVAVSH-----288
QY 305 YFGNGYGVKVG-----MRYSDRDADS-----NYAFAGSKLGMKTPAGRPCNCNTADDKAC 353
DB 289 ---QGDWQGVGQSDTYLOYEENTWKSREMSIDNTVF---KSTLVAPIGEH-----332
QY 354 AVGLGTEIKQKALAFDASYSRPPRLGNTANEFVIGADYNFRSTNEQOQRTTLYARGGLAL 413
DB 333 MLSFGVEGKHESLE-----DKTSNKI-----SSRTHISNTQW-----AGFIE 369
QY 414 NEFRSIPQVDLIANARKVGRYSHTVATENLDFEGLYKGSFTHP-----ADGLSLIGG 466
DB 370 DEWALAEQFRLTFGGR-----LDHDKNYG-SHFSPRVYGVWMLDLPLWTVKG 414
QY 467 GRLGHYKIESG-ECKTLHKASKT--KFTGYAGAVY---DLNNDNLSLYLSLSQLYTPQNL 520
DB 415 G-----VSTGFRAQPLREVTPDWQVSG-GGNYIGNPDLPQETSINKELSLMTSTSGSL 467
QY 521 DADGKLLKPRQNOFEVGYKGSYMDRLNARVSFYRMKDKNAAPLN-----PNN---KK 572
DB 468 AAS-----LTAFHNDFKDKITRVACPANICTAGPNQWGTATP 503
QY 573 TRYAALGRVMEGVETEISGAVTPKQIHAGYSYLHSQIKTASNSRDDGIFLLMPKHSAN 632
DB 504 TYRVNIDEAETYGAEATLSLPITESVELSSYTYTHSEQKSGNFAGRP--LLQLPKHLFN 561
QY 633 LWTYQVTPELTIGGVN-----AMSGITSSAGMHAGGYATFDMAAAYRFTPKLKLQIN 686
DB 562 ANLSWQTTDLNLSWANLNRYKQEMQPEGGASNDDFIAPSYTFIDTGVYALTDTATIKAA 621
QY 687 ADNIFNRHY-YARVGGANTFNIPGSERTW 714
DB 622 VYNLFDQEVNVAEYG-----YVEDGRRYW 645

RESULT 11

US-09-098-082-6
; Sequence 6, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

```

; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrpA amino acid sequence
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Vibrio Cholerae
;
US-09-098-082-6

```

```

Query Match          5.6%; Score 212; DB 3; Length 652;
Best Local Similarity 19.6%; Pred. No. 7,4e-11;
Matches 147; Conservative 99; Mismatches 299; Indels 204; Gaps 32;

```

```

QY 57 SPAVTVGKIPAS-----LRIPQSVSIITNOQVKNVDTF 93
DB 10 SLSTVGLMFSASAFQADATKTDETVWTAAGYAQVYQNPASISYRSRDLSERYRDV 69
QY 94 DQARTPGIRVLSNDGAS-SYARGYESEYNIDGLPQOMOST-----NGTLP 142
DB 70 TDALKSVPEVTVAGGDDTIDISIRGSGSNYTLTLVDGKQTSRQTRPNSDGPGEIGQWLP 129
QY 143 NLFAFDRVEVWRGSGGLFDSGEMGGLVNLVRKPTKAFQGHAAAGFGTHKQYKAADV 202
DB 130 PLQATIRIVIRPMSITLYSDAIGVINITKDOQOMGQVLSLVVQENNAASGDEQS 189
QY 203 GSINSDGVRGVRMAQTVGASPPRAEKNNH-----ETFYAADWDINDP--TVLGAG 253
DB 190 ANFFVTGPISDALSLQVGGTQTRDEDELEHGYGDKSLRSLTSLKLVQLNPLDQLLEAG 249
QY 254 YLYQGH-----LAPYNGLPADANNKLPSPQVFPVGAQDMNKRKM-----SHVFPADLKH 304
DB 250 VSAQDENNVKSAQSSGCGRTCSN-----TD-NQYRRHVAVSH-----288
QY 305 YFNGNGYGVKVG-----MRYSDRDADS-----NYAFAGSKLGMKTPAGRGCTADDKAC 353
DB 289 ---QGMQGVGSDTYLQVHEENTNKRMSIDTVF--KSTIVAPIGE-----332
QY 354 AVGLGTEIKOKALAFASYSRPFRLGNTANEFVIGADYNNFRSTNDRGRTTLVARGGLAL 413
DB 333 MLSFVGEGKHESLE-----DKTSNKL-----SSRTHISNQQW-----AGFIE 369
QY 414 NERSSTLPQVDLNNARKGVGRGSHYATENTLDFGIYKSTFHP-----ADGLSLIGG 466
DB 370 DEVALAEQFLTFGR-----LDHDKNYG-SHFSPRVYGVWNLDPMLTVKG 414
QY 467 GRGHYKIESG-EGKTLHKASKT--KFTGYAGAVY--DLNDNNSLYLSLSQLYTPQINL 520
DB 415 G-----VSTGRFAPQLREVTTPMGOVSG--GNNIYGNPDLQPTSTINKEISLMTYSGSL 467
QY 521 DADGKLLPRQGNQFEVYKGYKSYMDRLNARVSFYRMKDKNAAPLN-----PNN--KK 572

```

```

DB 468 AAS-----LTAFFNDFDKITRVACPNICTAGPNQKATP 503
QY 573 TRYAALGKMEVGEVETETISGAVTPKQIHAGYSYLSQIKTASNSRDDGIFLLMPGHSAN 632
DB 504 TYRANIDEAETGAEATLLELPITTESVELSSYTHYSBQSGNFAGRP--LQLPFHLEN 561
QY 633 LMTYQVTPBELTIGGVN-----ANSGITSSAGMHAGYATFDAMAAYRFTPKLQIN 686
DB 562 ANLSWGTTRRLNSWANLVRGKEMQEGGASNDDFIAPSYTFIDTGVTAALDTATIKAA 621
QY 687 ADNIENRHY-YARVGCANTFNIPGSERTW 714
DB 622 VYNLFDQEVNVAEYQ-----YVEDGRRYW 645

```

```

RESULT 12
US-08-537-361E-4
; Sequence 4, Application US/08537361E
; Patent No. 6121037
; GENERAL INFORMATION:
; APPLICANT: Scio|||kovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Hefton, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6121037e1 Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,361E
; FILING DATE: 02-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6121037nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 791 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-537-361E-4

```

```

Query Match          4.9%; Score 186.5; DB 3; Length 791;
Best Local Similarity 19.4%; Pred. No. 2,4e-08;
Matches 172; Conservative 103; Mismatches 329; Indels 283; Gaps 40;

```

```

QY 14 AATVLAALSSVFAQATADLET-----VHIKQSRYNALVTEKNDYSSFAVTG 64
DB 10 AALIGSIFGNPFVFADEAATETTPVKAEVKAVRVKQRPAPAVR-----VNLNR 60
QY 65 KIPASIREIPQSVSIITNOQVKNVDTFDQALARKTPGLRVLSNDG-----RSVYA 117
DB 61 IKQEMTRKNDKLVRYSTVDGLSDRSHQGFARVGEGRVGVSIDGVNLPDSEENSLYA 120
QY 118 RGEYSESYNIDGLPAQOMOSINGTLPNLFAFDREYVWRGSGGLFDSGEMGGLVNLVRKR- 176

```


Db 576 VSGTPECTEDDAYYYRCS-----PYEKLDWQMKIDKARIRIGIELTGRLANDKXA 627

Qy 592 GAVTPKMOIHAGYSYLSHQ-----IKTASNSRBDGIFLLMPKHSANL 633

Db 628 SFPEBGMKLFGLSGYAKSKLSGNSLSTOPLKVIAGIDVESPSEKMGVSRLTYLGAKK 687

Qy 634 WTTTYQVTPPELTIGGVNAMSGITSSAGMAGYATPDMAAAYRFTPKLKQINADNIFNR 693

Db 688 VKDAQTYVENKMGWGTPLQKRVQDYPMLNKSAY-VFDMYGFYKPAKLLTLRAGVYVLFNR 746

Qy 694 HY-----YARVG-----GANTFNIGSEBRTWTANLRYSF 722

Db 747 KTTWDSLRGLYSYSTTNAVDRDCKGLDRYRAG--RNYAVSLEWKF 791

RESULT 14

US-08-326-670A-2

/ Sequence 2, Application US/08326670A

/ Patent No. 5698438

/ GENERAL INFORMATION:

/ APPLICANT: Sciojiljovic, Igor

/ APPLICANT: So, Magdalene

/ APPLICANT: Hwa, Vivian

/ APPLICANT: Hefiton, Fred

/ APPLICANT: Nasait, Xavier

/ TITLE OF INVENTION: A No. 5698438e1 Bacterial Hemoglobin Receptor Gene and Uses

/ NUMBER OF SEQUENCES: 2

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: Allegretti & Witcoff, Ltd.

/ STREET: 10 South Wacker Drive, Suite 3000

/ CITY: Chicago

/ STATE: Illinois

/ COUNTRY: USA

/ ZIP: 60606

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patentin Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/326,670A

/ FILING DATE: 18 OCT 1994

/ CLASSIFICATION: 536

/ ATTORNEY/AGENT INFORMATION:

/ NAME: No. 5698438nan, Kevin E

/ REGISTRATION NUMBER: 35,303

/ REFERENCE/DOCKET NUMBER: 94,784

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 312-715-1000

/ TELEFAX: 312-715-1234

/ TELEX: 810-221-8317

/ INFORMATION FOR SEQ ID NO: 2:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 792 amino acids

/ TYPE: amino acid

/ TOPOLOGY: linear

/ MOLECULE TYPE: protein

US-08-326-670A-2

Query Match 4.9%; Score 185; DB 1; Length 792;

Best Local Similarity 19.6%; Pred. No. 3.3e-08;

Matches 173; Conservative 101; Mismatches 335; Indels 274; Gaps 39;

Qy 14 AATVLAALSSVFAAQTADLET-----VHIKQORSTNATVTEKNGYSSFAVTYGT 64

Db 10 AALVGSIFGVPVPAADAEATETTPVKAIVKAVKGGKRNAPAAVER-----VNLNR 60

Qy 65 KIPASLREITPOSYIITMOQVDRNVDTFQDLARKTFGLRVLSNDG-----RSSVYA 117

Db 61 IKQEMIRDNKDLYRYSITDVLSDSGRHQKGFPAVRGVGNKNGVSIQSVNLDPDEENSLYA 120

Qy 118 RGYEYSEYNDIGLPAQMQSINGTLPLNLFADRYEVMKGPGLFPDSSGEMGCIYVNLVYKR- 176

Db 121 R---YGNFNSRLSIDPELVR-----NIDIVKADSFNTSGALGGVNYOTLG 167

Qy 177 -----PTKAF-----OGHA-----AAGFGT-----HKQY 195

Db 168 RDLILPERQPGVMKMGKGYSTRNRWNTLTGFGVSNDRVDAALLYSORGHETESACKRGY 227

Qy 196 KAEADVSGSLNSDGSVRGVMAQTVGASPRPAKNNRHETPYAAMDWDINPDTVLGAGYL 255

Db 228 PVGAGSGA-NIRGSARG-----IPDPGQ--HKYHSFLGKIAYQJNDNHRIGASLN 275

Qy 256 YQQRH-----LAPNGLPADANNK-----LPSLPQHYFVGA--DMKFFKN 294

Db 276 GQGGHYTYVEESYNILASVYREADVDYNNRRNTNLFEYMTPESDRLSMVADVDYQRTKYS 335

Qy 295 SHDVPADLKHYPGNGSGYKVMRYSRDDADSNYAFAGSKLGMKTPAGRGCTATDPDKACA 354

Db 336 A---VNYKGFPIEDSSLTTRNYNOKLDELY-----NRSMD-----369

Qy 355 VGLTEIKOKALAFDASYSRPRFGLGNTANEFYIGADYNNFRSTNEQRTTLYARGLALN 414

Db 370 ---TFKRITRLD---SHPIQLG-----GRHLSFKTFASRR 402

Qy 415 EPRSIPOVDLIANAR--KGVKGSHTVATENDDEGTYGKSTFHPADGLSLIGGRLGHY 472

Db 403 DFNENLRDDYFSGRVVRRRTSSIQHPVKTNN--YGFSLSDQIQWNDVPSRAGIRYDHT 459

Qy 473 KIESGE-GKTLHKASKTK-----FTGYAGAVDYDLN-----DNNSLYL--SLSQLY 514

Db 460 KMTPOELNECHACDKTPPAANTYKMGSGFVGLAOLNQMKGVIITIGRYRPNASERY 519

Qy 515 -----TPQTNLDAD-----GKLLK--PROGNOPEVGYKGYMDR--LNA 550

Db 520 FTYNHSGMWLPNPLKABRTTHTLTLQGRSEKGLDANLYQSNVRFNLFSEBOKLTTSG 579

Qy 551 RVG-----FYRKDKNAAPLNPNNKTRRYAA--LGKVMGEV-----TEISGAT 595

Db 580 DVSCOTMNYYYGMS-----NPYSEKLEWQMNIDKARIRIGIELTGRLANDKXASFPV 632

Qy 596 PKQOIHAGYSYLSHQ-----IKTASNSRBDGIFLLMPKHSANLMTY 637

Db 633 EGMKLFGLSGYAKSKLSGNSLSTOPLKVIAGIDVESPSEKMGVSRLTYLGAKKXKXA 692

Qy 638 QVTPELTIGGVNAMSGITSSAGMAGYATPDMAAAYRFTPKLKQINADNIFNRHY-- 695

Db 693 QYTVENKMGWGTPLQKRVQDYPMLNKSAY-VFDMYGFYKPAKLLTLRAGVYVFNKRYTT 751

Qy 696 -----YARVG-----GANTFNIGSEBRTWTANLRYSF 722

Db 752 WSLRGLYSYSTTNAVDRDCKGLDRYRAG--SRNYAVSLEWKF 792

RESULT 15

US-08-990-470A-2

/ Sequence 2, Application US/08990470A

/ Patent No. 6123942

/ GENERAL INFORMATION:

/ APPLICANT: Sciojiljovic, Igor

/ APPLICANT: So, Magdalene

/ APPLICANT: Hwa, Vivian

/ APPLICANT: Hefiton, Fred

/ APPLICANT: Nasait, Xavier

/ TITLE OF INVENTION: No. 6123942e1 Bacterial Hemoglobin Receptor

/ NUMBER OF SEQUENCES: 5

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: McDonnell Boehnen Hubbert & Bergthoff

/ STREET: 300 South Wacker Drive, 32nd Floor

/ CITY: Chicago

/ STATE: Illinois

/ COUNTRY: USA

/ ZIP: 60606

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,470A
; FILING DATE: 15-DEC-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6123942nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-990-470A-2

Query Match 4.8%; Score 183; DB 3; Length 792;
Best Local Similarity 19.6%; Pred. No. 5.1e-08;
Matches 173; Conservative 101; Mismatches 335; Indels 274; Gaps 39;

Qy 14 AATVLAALSSSVFAAQTADLET-----VHIKQORSYNAIVTEKNGDYSSFAVTVGT 64
Db 10 AALVGSIFGNVPFAADEAATETTPVKAEVKAVRVKQGNAPAAVER-----VNLNR 60
Qy 65 KIPASLREIPQSVSIITNOQVDRNVDTFDQLARKTGLRLVLSNDDG-----RSSUYA 117
Db 61 IKQEMIRDNKDLVRYSTDVGLSDSGRHQKGFVRVGEGRVGVSDIGVNLDPSENSLYA 120
Qy 118 RGVSEYSEYNIDGLPAQMQSINGTLPNLFAPDRVEMRGPGLFDSSGEMGIVNLVLRK- 176
Db 121 R---YGNFNSRLSIDPELVR-----NIDIVKGADSFNTSGALGGGVNYQTLOG 167
Qy 177 -----PTKAF-----QGHA-----AAGFGT-----HKQY 195
Db 168 RDLLLPERQFGVMKNGYSTRNRENTLFGVGNDRVDAAALYSQRGHETESAGKGY 227
Qy 196 KAEADVSGSLNDGSGVRGVMAQTGASPRPAEKNNRHETFYAADWDINPDTVLGAGYL 255
Db 228 PVEGAGSGA-NIRGSARG-----IPDPQ--HKYHSFLGKIAYQINDNHRIGASLN 275
Qy 256 YQORH-----LAPYNGLPADANNK-----LPSLPQHVFGA--DMNKFQON 294
Db 276 GQGHNYTVESYNLLASTWRADDVNRNRTNLFYEWTPESDRLSWKADVDYQTKVS 335
Qy 295 SHDVFADLKHFGNGGYGVGMRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACA 354
Db 336 A---VNYKGSFPIDESTLTRNYNQKLDLEIV-----NRSM-----369
Qy 355 VGLGTEIKOKALAFDASYRPRFLGNTANEFVIGADYNFRSTNEQGRITLYARGGLALN 414
Db 370 -----TRFKRITRLD---SHPLQLGG-----GHRLSFKTFASRR 402
Qy 415 EFRSIPQVDLIANAR--KGVRGVSHVTATENLDEFGLYKSTFHPADGLSLIGGGRGLGHY 472
Db 403 DENLNRRDYYFSGRVVRTSSIQHPVKTN---YGFSLSDQIQWNDVFPSSRAGIRYDHT 459
Qy 473 KIESGE-GKTLHKASKTK-----FTCYAGAVVDLN-----DNNSLYL--SLSQLY 514
Db 460 KWTQPQLNAECHACDKTPPAANTYKWSGFGVGLAQLAQAWRVGVYDITSGYRVPNASEVY 519
Qy 515 -----TQTNLDAD-----GKLLK-PRQGNQFEVYKGSYMDR---LNA 550
Db 520 FTYNHGSGNWLNPNLKABERTTTHTLSLQGRSEKGTLDANLYQSNRYNPLFSEQKLTTS 579
Qy 551 RVS-----FYMKDKNAAPLNPNKTRYAA--LGKRVMEGVE-----TEISGAVT 595
Db 580 DVSCTQMYYNYGWS-----NPYSEKLEWQWQIDKARIGLELTGELNVKDVKASFPV 632

```

```

Qy 596 PKWQHAGYSYLHSQ-----IKTASNRDDGIFLLMPKHSANLWTTY 637
Db 633 EGWKLFGSGYAKSKLGDNSLLSTQPLKVIAGIDYESPSEKMGVFSRLTYLGAKKVKDA 692
Qy 638 QVTPELTITGGGVNAMSGITSSAGMHAGGYATFDAMAAVRYFTPKLKLQINAQNIENRHY-- 695
Db 693 QYTVYENKMGWTPLOKKVKDYPWLNKSGAY-VFDMYGFYKVPVKNTLRAGVNVNFRKTYT 751
Qy 696 -----YARVG-----GANTFNIPGSERTWTANLRYSF 722
Db 752 WDSLRLGLYSYSTTNSVDRDGKGLDRYRAP--SRNYAVSLEWK 792

```

```

RESULT 16
US-08-817-707-2
; Sequence 2, Application US/08817707
; Patent No. 6277382
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-707-2

```

```

Query Match 4.8%; Score 183; DB 4; Length 792;
Best Local Similarity 19.6%; Pred. No. 5.1e-08;
Matches 173; Conservative 101; Mismatches 335; Indels 274; Gaps 39;

Qy 14 AATVLAALSSSVFAAQTADLET-----VHIKQORSYNAIVTEKNGDYSSFAVTVGT 64
Db 10 AALVGSIFGNVPFAADEAATETTPVKAEVKAVRVKQGNAPAAVER-----VNLNR 60
Qy 65 KIPASLREIPQSVSIITNOQVDRNVDTFDQLARKTGLRLVLSNDDG-----RSSUYA 117
Db 61 IKQEMIRDNKDLVRYSTDVGLSDSGRHQKGFVRVGEGRVGVSDIGVNLDPSENSLYA 120
Qy 118 RGVSEYSEYNIDGLPAQMQSINGTLPNLFAPDRVEMRGPGLFDSSGEMGIVNLVLRK- 176
Db 121 R---YGNFNSRLSIDPELVR-----NIDIVKGADSFNTSGALGGGVNYQTLOG 167

```


Db 587 THPKLLSANPEGAQTGTWTASLAIRKFNLENLEIGWGRVYQKATGSILAAQKDRDG 646
 Qy 663 -----AGVATFDAMAARFTPK--LKQINADNIFNRHYIYARVGGANTFNIPGSERTW 714
 Db 647 KLENVVRQGFVNDVFNANKPLGKOTLVNLSVNNVDFKFIY-----PHSQR-W 694
 Qy 715 TANL 718
 Db 695 TINTL 698

RESULT 18
 PCT-US95-06994-7
 ; Sequence 7, Application PC/TUS9506994
 ; GENERAL INFORMATION:
 ; APPLICANT: Children's Hospital & Medical Center
 ; APPLICANT: Washington State University Research Foundation
 ; APPLICANT: TARR, PHILLIP I
 ; APPLICANT: BILGE, SIMA S
 ; APPLICANT: BESSER, THOMAS E
 ; APPLICANT: VARY JR, JAMES C
 ; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 ; STREET: SUITE 2800, 1420 FIFTH AVENUE
 ; CITY: SEATTLE
 ; STATE: WASHINGTON
 ; COUNTRY: USA
 ; ZIP: WA 98101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/06994
 ; FILING DATE: 07-JUN-95
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/265,714
 ; FILING DATE: 24-JUN-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BRODERICK, THOMAS F
 ; REGISTRATION NUMBER: 31,332
 ; REFERENCE/DOCKET NUMBER: CHOR-18591
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 682 8100
 ; TELEFAX: (206) 224 0779
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 718 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; DESCRIPTION: Vibrio cholerae ICGA amino acid sequence,
 ; DESCRIPTION: wherein Xaa" residues represent gaps
 ; DESCRIPTION: introduced to facilitate best alignment with
 ; DESCRIPTION: SEQ ID NO:6.
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: VIBRIO CHOLERAEE

PCT-US95-06994-7
 Query Match 4.8%; Score 180.5; DB 5; Length 718;
 Best Local Similarity 18.8%; Pred. No. 7.4e-08;
 Matches 146; Conservative 92; Mismatches 285; Indels 255; Gaps 31;
 Qy 57 SFATVVGTKIPAS-----LREIPQSVSLITNQVQKDRVD 91
 Db 10 SUSVTGLMFSASXAXFAQDATKTDETVMVTAAGYAQVIONAPASISVISEDLESYYR 69

Qy 92 TFDQARKTPGLRLVLSNDDGRS-----SVYARGYSEYNIDGLPAQMOSI----- 137
 Db 70 DVTDAKSLVPGVTYXXXTGGDDTTDISIRMGSGNYTILILVDGKRQTSRQTRPNSDGPGE 129
 Qy 138 NGTLPNLPADRVEMRGPSGLFDSSGEMGIVNLVRKPTKAFQCHA-----AA 187
 Db 130 QGWLPPLOAERIEVIRGPMSTLYGSDAIGVINIITRKQOQOQSGNVQLSTVVQENRAS 189
 Qy 188 GFGTHQVKAEDVSGSLNSDGSVRGRVMAQTVGASPRPAE-----KNNRHETFYAAADW 242
 Db 190 G-----DEQSANFVTGPL--SDXALSLOVYGQTTORDEDEIEHGYGDKSLRSLT--SKLNY 242
 Qy 243 DINPD--TVLGAGLYQQRH-----LAPYNGLPADANNKLPSPLOHVFVG-- 285
 Db 243 QLNPDHQLQLEAGVSAQDRENNVGKSAQSGCXXXXXXRGTCSTNDNQYR--NHVAVSH 300
 Qy 286 -ADNKKPMNSHDVPADLKHVFGNGYGVGMRYSDRDADS----- 325
 Db 301 QGDWQ-----GVXGSDTYLQYEENTNKSREXXXXXXKXXXXXXX 343
 Qy 326 -----NYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKOKALAFADASYSRPRLG 379
 Db 344 XXMSIDNTVP---KSTLVAPIGEH-----MLSGVEGKHESLE-----D 379
 Qy 380 NTANEFVIGADYNRFRSTNEQRTTLYARGGLALNEFRSIPQVDLIANARKGVRGYSHTV 439
 Db 380 KTSNKI-----SSRTHISNTQW-----AGFIEDWALAEQPRLTFGGR----- 417
 Qy 440 ATENLDEFGIYKSTFHP-----ADGLSLIGGRIGHYKIESGEGKTLHKASKTKFT- 491
 Db 418 ----LDHDKNYG-SHFSRVYGVWNLDPMTVKG-----VSTG-----FRAPQLREVT 461
 Qy 492 -----GVAGAVY---DLNDNNSLYLSQLYTPOTNLDDAGLKLKPRQGNQFEVGYK 540
 Db 462 PDWQVSGGGXXNIYGNPDLOPETSINKLSLWYSTGSLAAS----- 504
 Qy 541 GSYMDDRNLNARVSFYRMKDKNAAAPLN-----PNN-----KKTRYAALGRVMEGVETRISG 592
 Db 505 -----LTAPHNDFKDKITRVACPANICTAGPNQMGATPYRVNIDEAETYGAEATLSL 557
 Qy 593 AVTPKQIHAGYSYLHSQIKTASNSRDDGIFLLMPKHSANLWTTQVTPELTIGGGYNAM 652
 Db 558 PITSEVELSSSYTYTHSEQSGNFAGRP--LLQLPKHLFNANLSWQTTDLNLSWANLNYR 615
 Qy 653 SGITSSAGMEAG-----GVATPDMAAAYRFTPKLKLQINADNIFNR 693
 Db 616 GKXXEMQPEGASNDDFIAPXXXXXXSYTFIDTGTALTDTATIKAAVYNLFDQ 673

RESULT 19
 US-08-537-361E-2
 ; Sequence 2, Application US/08537361E
 ; Patent No. 6121037
 ; GENERAL INFORMATION:
 ; APPLICANT: Stojiljkovic, Igor
 ; APPLICANT: So, Magdalene
 ; APPLICANT: Hwa, Vivian
 ; APPLICANT: Heffron, Fred
 ; APPLICANT: Nassif, Xavier
 ; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 ; STREET: 300 South Wacker Drive, 32nd Floor
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 792 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-361E-2

Query Match 4.8%; Score 180; DB 3; Length 792;

Best Local Similarity 19.4%; Pred. No. 9.7e-08;
Matches 172; Conservative 104; Mismatches 328; Indels 284; Gaps 42;

QY 14 AATVLAALSSVFAAQTADLETVHI-----KQSRSYNAIVTEKNGDYSSFAVTG 64
DB 10 AALVGSIFGNPVPFAADEATETTPVKAAYKAVKQGRNAPAAVER-----VNLNR 60
QY 65 KIPASIREIPQSVSIITNOQVDRNVDTPDQARKTGRLVLSNDG-----RSSVYA 117
DB 61 IKQEMIRDMKDLVRYSTVDGLSDSGRHQKFAVRGEGNRVGVSIDGVNLPDSEENSLYA 120
QY 118 RGYEYSEVNIIDGLPAQMOSINGTLPLPAFDREYVWRGSPGLFDSGEMGIVNLYVRK- 176
DB 121 R---YGNFNSRLSIDPELVR-----NIDIVKADSFNTSGALGGVNYOTLOG 167
QY 177 -----PTKAF-----QGHAF-----AAGFGT-----HKQY 195
DB 168 RDLLEPERFGVWMMKGYSTRNREWTNTLGFVGSNDRVDAALLYSQRGHETESAKRGY 227
QY 196 KAEADVSGSLSDSGYRGVMAQTVGASPPRAEKNRHETFAADWDINDPTVLGAGY 255
DB 228 PYEGAGSGA-NIRGASRG-----IDPPSQ--HKYHSFLGKIAYQINDNHRIGASIN 275
QY 256 YQORH-----LAPYNGLPADANK-----LPSLPQHVFYGA--DMNKFKKN 294
DB 276 GQGGHNYVEESYNLLASTYWRADVDNRRRTNLFYEWTPESDRLSMVKADVDYQTKVYS 335
QY 295 SHDVFADLKHYFGNGGYGKVMRYSDRDADSNYAFASGKLGMKTPAGRCQNTAADKACA 354
DB 336 A-----VNYGSPFIEDSSTLTIRYNNOKDDELY-----NRSM----- 369
QY 355 VGLGTEIKOKALAFDASYSRPFELGNTANEVIGADYNFRSTNEQRTTLVYARGGIALN 414
DB 370 -----TFPKRITRLD---SHPLOLG-----GRHRLSKPTASRR 402
QY 415 EFRSTPQVULLINAR--KAVRGSHTVATENLDEFGTYGKSTHPADGSLIGGGLGHY 472
DB 403 DEENLRDYYFSGRVRTTSSIOHPVKTTN--YGFSLSDQIOMNDVSSRAGIYYDHT 459
QY 473 KIESGE-GKTLHKASKTK-----FTGYAGAVYDLN-----DNNSLYL--SLSOLY 514
DB 460 KMTPEQLMECHACDCTTPRAANTYKMGSGFVGLAALNQAMKRGYITSGIRYPNASEVY 519
QY 515 -----TQQTNLAD-----GKLLK--PRQGNQFEVGYKGSYMDR--LNA 550
DB 520 FTYNHSGSNMLPMPNLKABERTTHTLSLOGRSKGLDANLYQSNRYNRLSEBOKLTTSG 579
QY 551 RVS-----FYRKDKKAAAPLNPNNKTRRYAA--LGKVMGEV-----TEIGAYT 595
DB 580 DVSCTQMNYYYGKCS-----NPYSEKLEWQNONIDKIRIGIELTGLRINDVKASFPV 632

QY 596 PFKQIHAGYSYHSQIKTASNRDGIFFLMP-----KHSANLW-----TTY---- 637
DB 633 EEWKLFJGSLGYAKSKL-----SGDNSLFTQPLKVIAGIDYSPSEKMWGFSRLTYLGAK 687
QY 638 -----QVPELITGGVNMAMSGITSSAGNHAGGYATFDMAAAYRFPKJLQIOMDINFN 692
DB 688 KYKDAQYTYENKGMGTPIQKKVKDYPMLNKSAV--VFDYGYGFKYVKNTLLRAGYVNVN 746
QY 693 RHY-----YARVG-----GANTFNIPEGSERTWTANLRYSF 722
DB 747 RYTTWDSLRGLYSYSTNSVDRDQKGLDRYAP--SRNYAVSLERKF 792

RESULT 20

US-08-537-361E-6
Sequence 6, Application US/08537361E
Patent No. 6121037

GENERAL INFORMATION:
APPLICANT: Stojiljkovic, Igor
APPLICANT: So, Magdalene
APPLICANT: Hwa, Vivian
APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037e1 Bacterial Hemoglobin Receptor
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 792 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-361E-6

Query Match 4.7%; Score 179; DB 3; Length 792;

Best Local Similarity 19.8%; Pred. No. 1.2e-07;
Matches 175; Conservative 96; Mismatches 338; Indels 274; Gaps 40;

QY 14 AATVLAALSSVFAAQTADLET-----VHIKQSRSYNAIVTEKNGDYSSFAVTG 64
DB 10 AALVGSIFGNPVPFAADEATETTPVKAAYKAVKQGRNAPAAVER-----VNLNR 60
QY 65 KIPASIREIPQSVSIITNOQVDRNVDTPDQARKTGRLVLSNDG-----RSSVYA 117
DB 61 IKQEMIRDMKDLVRYSTVDGLSDSGRHQKFAVRGEGNRVGVSIDGVNLPDSEENSLYA 120
QY 118 RGYEYSEVNIIDGLPAQMOSINGTLPLPAFDREYVWRGSPGLFDSGEMGIVNLYVRK- 176
DB 121 R---YGNFNSRLSIDPELVR-----NIDIVKADSFNTSGALGGVNYOTLOG 167

177	Qy	PTKAF-----QGHA-----AAGFGT-----HKQY 195
168	Db	RDLLPRLPQFGVMKNGYSTRENTWTNLTGFGVSNDRDVAALLYSORRGHETESACKRGY 227
196	Qy	KAEADVSGSLNSDGSVGRVMAQTVGASPPAEKNNRHETFYAAADWDINPDTVLGAGYL 255
228	Db	PVEGAGSGA-NIRGSARG-----IPDPSQ--HKYHSFLGKIAYQINDHRIAGSLN 275
256	Qy	YQQRH-----LAPYNGLPADAANK-----LPSLPQHVFVGA--DMWKFKN 294
276	Db	GOQHNYTVEESYNLLASVYREADDVNRRNTNLFEWTPESDRLSNWADVDYQKTKVS 335
295	Qy	SHDVFADLKHVFGNGYGVKMYSDRDASNVAFAGSKLGMKTPAGRPCNCTADDKACA 354
336	Db	A-----VNYKSGFPEDSSTTRNTNQKDOEII-----NRSM-----369
355	Qy	VGLGTEIKQKALAPDASYSRPFRLGNTAN--EFVIGADYNRPSTNEQGRITTLTYARGGLA 412
370	Db	---TRFKRITTLRLD---SHPLQLGGGRHRLSKFTFASRRDPENLRD---YYPSGRV 418
413	Qy	LNFRSPITPOVDLLANARKGVRGYSHVTATENLDEFGIYKGSTTHPADGLSLGCGRLGHY 472
419	Db	VRTTSSI-----QHPVKYTN-----YGFSLSDQIQWINDVFSRRAGIRYDHT 459
473	Qy	KIESGE-GKTLHKASKTK-----FTGVAGAVYDLN-----DNNSLYL--SLSQLY 514
460	Db	KWTFQELNAECHACDKTPPAANTYKKGSHGFVGLAAQLNQAWRVGYDITSGYRVPNASEVY 519
515	Qy	-----TPQTNLDAD-----OKLJK-PRQGNQFVGVKGYGSMDDR---LNA 550
520	Db	FTYNHGSGNLPNPNLKAERSTTHTLSQLGRSEKGTLDANLYQSNRYNFLSEBQKLTTSG 579
551	Qy	RVS-----FVRMKDKNAAPLNNKKTYAA--LGKRVMGEV-----TEISGAVT 595
580	Db	DVSCQTQWNYIGMCS-----NPYSEKLEWQNDIDKARIGELTGRNLVDVKASFVP 632
596	Qy	PKWQIHAGYSYLHSQ-----IKTASNSRDDGIFLLMPKHSANLWTTY 637
633	Db	EGWKLFGSLGYAKSKLSGDSNLLSTQPLKVIAIGIDYESPEKGVFSRLTYLCAKKVKDA 692
638	Qy	QVTPELTIGGVNAMSGITSSAGMHAGGYATPDMAAYRFTPKLKLQINADINRHY-- 695
693	Db	QYTVYENKMGWTPLOKKVQDYDPMUNKSAY-VFDMWGYFVKPNKLTURAGVYNVFNKRYTT 751
696	Qy	-----YARVG-----GANTFNIPGSERTWTANLRYSF 722
752	Db	WDSLGLGVSYSTTNSVDRDCKGLDVRAP--SRNVAUSLEWKF 792

```

Db 577 VS--HSKRFYDTHKDK-LLSANPERGAOVGRWTWASLAVRFKNPULIEGMRGRTYQKAV 633
Qy 654 GITSAGMH-----AGGYATPDMAAAVAFPK--LKLQINADNIENRHYAVRVS 701
Db 634 GSILAAQGRDROKLENNVRQGGVNDVAFNMKPLKQDITLNVLSVNNVDFDKRY----- 688
Qy 702 ANTFNIPGSERTWTANL 718
Db 689 -----PHSOR-WTNTL 698

RESULT 22
US-08-613-009A-18
; Sequence 18, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loommore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-613-009A-18

Query Match: 4.6%; Score 173; DB 3; Length 915;
Best Local Similarity 21.4%; Pred. No. 5.5e-07;
Matches 172; Conservative 99; Mismatches 292; Indels 242; Gaps 43;
Qy 3 QPMVSRIN--MTAAVIALSSVFA--AQTADLETYHIKQSRYNALVTEKNDYSS 57
Db 2 QOQHFRLLIILICISMTALPAVAENVQAOQKQDITGVAKKQKTRDNEVTG---- 57
Qy 58 FAVTVGTIKIPALREIPGVSIITNQOVDRVNDTFDQAKRTPGIKRVLSNDDGRSVYA 117
Db 58 -----LGKIVKTAADTLISKQVID-----IRDLTRYDPGIAVVEQGGAGSSGYS 100
Qy 118 -RGYEYSEYN--IDGLPAQMOG-----INGTLP-----NLFAPDR--VEVMKGPFG 158
Db 101 IRGMKNNRSLVVDGL-AQIQSTYAOALGGRITGSSGAINIEIENVKAVAFISGNS 159
Qy 159 LFDSSGEMGIVNLVKRPDTKAFQGHAAAG-PGTHKQYAEADVSGSLNSDGSVGRGVNA 217

```

```

Db 160 VEGSGALAGSV-----AFQTKTADVDVIGEGQWGIQOSTAYSGKNNRGLTQSLALA 210
Qy 218 QTVGASPRPAEKNNRHETFYAAADMDINPDVLGGLYVQGNHLAY----- 264
Db 211 GRIGAEALLITGRHA-----GEIRAHAAGRG-VQSFNRLAVDDGSKYAFIVEE 262
Qy 265 -----NGLPADANNKLPSPHVEFGADW-----NKEPMN--SHDV----- 298
Db 263 ECKNG-----GHEKCKANPKPDVGEDKQTVGTROVTGPNRFLADPLSESSWMLRPGF 318
Qy 299 -PADLKHYPNGGYGKVKMRYSGRDA-----DSNVAPAGSKLGMKTPAGE--P 343
Db 319 REENKHHYIG-GILERTQOTFDTRODTWTPAFLTKAVFDANQKQAGSLRGNGKXAGNHKYG 377
Qy 344 GCNTADDKACAVG--LGTEIKOKALAFDASYSRPFELGTANEFVIGALYNNRFRSTNEG 401
Db 378 GLFTSGENNAPVGAEGYGVTFYDETHTKSRVGLGYTTNADKO--TWADYARLS----- 429
Qy 402 RTLLVARGGLAL-NEFRSIPQVDLIANARKGVRSHTVATENLDEFG-----LYGKSTH 456
Db 430 ----YDRQIGLDNHQ--QTHCSAD--GSDKYCRPADKRFSSYKSDRVYIGRS----- 476
Qy 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKYF-----TGYAGAVYDINDNN 504
Db 477 -----HKLQAAFKKSFDTAKIRHNLVNLGYDRFGSNLRHOD 514
Qy 505 SLVLSISQLYTPQTNLDADGKLLK--RGNOPEVGYKKSVMYDDRLNANVSFPRMCKXA 562
Db 515 YYYQSANRAVYSSKTPPQNNKKTSPNGRKNPYWSI-----GRNVVTRQICLGNNNT 568
Qy 563 AAPLNP--NKKTRYAALGKRWEGVETISGAVTPKW-QIHAGYSY-----LHS---QIK 612
Db 569 YDCTPRSLNGSKSYAAVANDNRLG-----RMDVGAGLRYDVRSHSDGYS 617
Qy 613 YASN--SRDDGIFLIMPRHSANLWTTVQVPELITGGVNMAGITTSAGMHAGYATF 669
Db 618 TQTHRLTSWNAQIVL--KPADWLDTYR-----TSTGRLLPFA-- 654
Qy 670 DMAAARFPTKXK-LQINADNIENR 693
Db 655 -EMYGWRSQDKIKAVXIDPKSFK 678

RESULT 23
US-08-778-570B-24
; Sequence 24, Application US/08778570B
; Patent No. 6437096
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loommore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 03-JAN-1997

```

```
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-664
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-778-570B-24

Query Match 4.6%; Score 173; DB 4; Length 915;
Best Local Similarity 21.4%; Pred. No. 5.5e-07;
Matches 172; Conservative 99; Mismatches 292; Indels 242; Gaps 43;

QY 3 QPMSVFRIN---MTAATVLAALSSSVFA--AQADLETVHIKQORSYNAIVTEKNGDYSS 57
DB 2 QOQHLFRNLCLSLMTALPAYAENVQAGQAEKQLDTIQVAKKQKTRRDNEVTG--- 57
QY 58 FAVTVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQARKTPGLRLVLSNDDGRSSVYA 117
DB 58 -----LGKLVKTADTLSKEQVLD-----IRDLTRYDPGIAVVEQGRGASSGYS 100
QY 118 -RGYEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAPDR---VEVMRGPSS 158
DB 101 IRGMDKNRVLSTVDGL-AQIQSYTAQAALGGTRTAGSSGAINEIYENKAVEISKGSNS 159
QY 159 LFDSSGEMGGIVNLVRKRPKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVGRVMA 217
DB 160 VEOGSGALAGSV-----AFQTKTADDVIGEGRWGIQSKTAYSGKNRGLTQSLALA 210
QY 218 QTVGASPRAEKKNRHETFYAADWDINDPTVLGAGLYQOHLAPY----- 264
DB 211 GRIGGAELIIRTRHA-----GEIRAHAAGRG-VQSFNRLAPVDDGSKYAYFIVEE 262
QY 265 ---NGLPADANNKLPSPQHVFVGADW-----NKFKN--SHDV----- 298
DB 263 ECKNG-----GHEKCKANPKDVUGEDKQRTVTRDTYGTGNRFLADPLSYESRWLPQGF 318
QY 299 -PADLKHYFGNGGYGVGMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
DB 319 RFENKRHYIG-GILERTQOTFTDRMTVPAFLTKAVFDANQOAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDDKACAVG--LGTEIKOKALAFDASYSRPFLGNTANEFIGADYNFRSTNEQG 401
DB 378 GLFTSGENNAPVGAEGYGTGVFYDETHKSRYLEVYVYTNADKD--TWADYARLS----- 429
QY 402 RTTLVARGGLAL-NEPRSIPOVDLIANARKVGRGYSHTVATENLDEFG---IYKSTFH 456
DB 430 ----YDRQIGILDNIHQ---QTHCSAD-----GSDKYCRPSADRFPSYKSDRVIYES--- 476
QY 457 PADGLSLIGGRGLGHYKIESGEGKTLHKASKTKF-----TGAGAVYDLNDNN 504
DB 477 -----HKLQAFAFKKSFTDAKIRHNLNVNLGYDRFGSNLRHQD 514
QY 505 SLVLSLSQLYTQTNLDADGKLLK--ROGNQPEVGYKGSYMDRLNARVSYRMRKDKNA 562
DB 515 VYQSANRAYSKTTPQNNKGKTSNPGNEKNPYWVSI-----GRGNVTRQICLFGNNT 568
QY 563 AAPLNP--NKTRYAALCKRMVEGVETEISGAVTPKW-OIHAGYSY-----LHS---QIK 612
DB 569 YTDCTPRSRINGKSYAAVRDNRVLG-----RWADVAGRLRYDRSTHSDGSGVS 617
QY 613 TASN---SRDDGIFLLMPKHSANLMTTYQVTPLETITGGGVNAMSGITTSAGMHAGGYATF 669
DB 618 TGTHTRLSNAGIVL---KPADWLDLTYR-----TSTGPRLPSPFA-- 654
QY 670 DAMAAYRFTPKLK-LQINADNLFNR 693
```

```
DB 655 -EMYGWRSGDKIKAVKIDPEKSPNK 678

RESULT 24
US-09-059-584-24
; Sequence 24, Application US/09059584
; Patent No. 6440701
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/059,584
; APPLICATION NUMBER: US/09/059,584
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-059-584-24

Query Match 4.6%; Score 173; DB 4; Length 915;
Best Local Similarity 21.4%; Pred. No. 5.5e-07;
Matches 172; Conservative 99; Mismatches 292; Indels 242; Gaps 43;

QY 3 QPMSVFRIN---MTAATVLAALSSSVFA--AQADLETVHIKQORSYNAIVTEKNGDYSS 57
DB 2 QOQHLFRNLCLSLMTALPAYAENVQAGQAEKQLDTIQVAKKQKTRRDNEVTG--- 57
QY 58 FAVTVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQARKTPGLRLVLSNDDGRSSVYA 117
DB 58 -----LGKLVKTADTLSKEQVLD-----IRDLTRYDPGIAVVEQGRGASSGYS 100
QY 118 -RGYEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAPDR---VEVMRGPSS 158
DB 101 IRGMDKNRVLSTVDGL-AQIQSYTAQAALGGTRTAGSSGAINEIYENKAVEISKGSNS 159
QY 159 LFDSSGEMGGIVNLVRKRPKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVGRVMA 217
DB 160 VEOGSGALAGSV-----AFQTKTADDVIGEGRWGIQSKTAYSGKNRGLTQSLALA 210
```



```

Db 618 TGHRTLSWAGLVL---KPADWLDLTYR-----TSTGRPLPSFA-- 654
QY 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWRSGBKIKAVKIDPEKSFNK 678

RESULT 26
US-08-363-124A-2
; Sequence 2, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varot Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-363-124A-2

Query Match 4.6%; Score 172; DB 2; Length 915;
Best Local Similarity 21.4%; Pred No. 6,9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

QY 3 QPMSVFRIN---MTAATVLAALSSSVA--AQTALETVHIKQRSYNAIVTEKNGDYSS 57
Db 2 QOQHLFRNLCLSLMTALPAVAENVQAGQEKQDITQVAKKQKTRRDNEVTG--- 57
QY 58 FAVTVGTKIPASLREIPQSVSLITNQVYKDRNVDFDQLARTPTGLRVLSNDDGSSUYA 117
Db 58 -----LGLKLVKTADTSLSEKQVLD-----IRDLTRYDPGIVAVVEQGRGASSGYS 100
QY 118 -RGVEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAFDR---VEVMRGPSP 158
Db 101 IRGMDKNRSLTVDDL-AQIQSYTAQAALGGTRTAGSSGAINIEIYENVKVAIEISKGSNS 159

```

```

QY 159 LFDSSGEMGGIWNLVKRKPTKAFQGHAAAG-FGTHKQYKABADVSGSLNSDGSVGRVRWA 217
Db 160 VEOGSGLAGSV-----AFQTKTADDVIGEGROWGIOQSKTAYSGKNRGLTQSIALA 210
QY 218 QTVGASPRPAEKNRHHETFYAAADWDINPDIVLGAGVLYQOORHLAPY----- 264
Db 211 GRIGGAELLIRTORHA-----GEIRAHAAGRG-VQSNRNLAPVDDGSKYAYFIVEE 262
QY 265 ---NGLPADANNKLPSLPQHVFVGADW-----NKFQKN---SHDV----- 298
Db 263 ECKNG---GHEKCKANPKDDVGVGDKRQTVSTRDYTCNPRFLADPLSYESRSWLPGRPF 318
QY 299 -PADLKHFGNGGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRMTVPAFLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEQG 401
Db 378 GLFTSGENNAVPVGAEGYGTGVFVDETHKRSRYGLEBYVTNADKD---TWADYARLS--- 429
QY 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKGVRGYSHTVATENLDEFG---IYCKSTFH 456
Db 430 ---YDRQIGLDNHFQ---QTHCSAD---GSDKYCRPSADKPFYSYKSDRVIYGES--- 476
QY 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKTKP-----TCYAGAVVDLNDNN 504
Db 477 -----HKLQAAFKKSPDTAKIRHNLSVNLGYDRFGSNLRHQD 514
QY 505 SLVLSLSQLYTPQTNLDADGKLLKP--ROGNQFVGVKGSYMDORLNARVSFYRMKDQNA 562
Db 515 YYYOSANRAYSLKTPPONNGKKTSPNGREKNPYWWSI-----GRGNVVTQICLFGNNT 568
QY 563 AAPLNP--NKKTRYAALGKRVMEGVETEISGAVTPKW-QIHAGYSY-----LHS---QIK 612
Db 569 YDCTPRSINGKSYAAVRDNVRLG-----RWADVGAGLRVDYRSTHSDGGSVS 617
QY 613 TASN---SRDDGIFLLMPKHSANLWTTVQVTPELTIGGVNAMSGITSSAGMHAGGAVTF 669
Db 618 TGTHRTLSWAGLVL---KPADWLDLTYR-----TSTGRPLPSFA-- 654
QY 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWRSGBKIKAVKIDPEKSFNK 678

```

RESULT 27

```

US-08-478-435-96
; Sequence 96, Application US/08478435
; Patent No. 5923232
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murlin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```


Db 58 -----LGLKLVKTADTLKSKEQVLD-----IRDLTRYDPGIAVVEQGRGASSGS 100
Qy 118 -RGYEYSEYN-IDGLPAQOMS-----INGTLP-----NLFAPDR---VEVMRGPSSG 158
Db 101 IRGMDKNRVSLTVDDL-AQIQSYTAQAALGGTRTAGSSGAINIEIYENVKAVEISKGSNS 159
Qy 159 LFDSSGEMGGIVNLVRKRTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVGRVMA 217
Db 160 VEQSGGALAGSV-----AFQTKTADDVIGEGRGWGIQSKTAYSGKNRGLTQSIALA 210
Qy 218 QTVGASPRPAEKNNRHETFYAAADWDINDPTVLGAGLYQOORHLAPY----- 264
Db 211 GRIGGAELLITGRHA-----GEIRAHEAAGRG-VQSFNRLAPVDVDSKYAYFIVEE 262
Qy 265 ---NGLPADANNKLPSPHOFVFGADW-----NKFKNM---SHDV----- 298
Db 263 ECKNG-----GHEKCKANPKKDVVGEDKQTVSTRDYTGPNRFLADPLSYESRSLPRPGP 318
Qy 299 -FADLKHYFGNGGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRMTVPAPLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
Qy 344 GCNTADKACAVG--LGTEIKQKALAFDASYSRPFLGNTANEFVIGADYNNFRSTNEQG 401
Db 378 GLFTSGENNAPVGAEGYGVDFETHKSRYLEVYVYNADKD--TWADYARLS----- 429
Qy 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKGVGYSHTVATENLDEFG-----IYKSTFFH 456
Db 430 ----YDRQIGLDNHFO---QTHCSAD---GSDKCYRPSADKPPSYKSDRVIYGES--- 476
Qy 457 PADGLSLGGRLGHYKIESGEGKTLHKASKTKF-----TGAGAVYDLDNNN 504
Db 477 -----HKLQAFKKSFTAKIRHNLVSNLGYDFRGSNLRHQD 514
Qy 505 SLYLSLSQLYTPQTNLDADGKLLKP--ROGNOFEYGYKGSYMDRLNARVSFYRMKDKN 562
Db 515 YYQSANRAYSLKTPQNGKKTSPNGREKNYVWSI-----GRGNVTVROIQLFGNNT 568
Qy 563 AAPLNP--NKTRYAALGKRYMEGVETEISGAVTPKW-QIHAGYSY-----LHS---QIK 612
Db 569 YTDCTPRSINGSYAAVRDNRVLG-----RWADVAGLRXYDRSTHSDGGSVS 617
Qy 613 TASN---SRDDGIFILMPKHSANLITVYQVTPELTIGGVNAMSGITSSAGNHAGGYATF 669
Db 618 TGTHRTLSSNAGIVL---KPADWLDLTVR-----TSTGRLPSFA-- 654
Qy 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWSGDKIKAVKIDPEKSFNK 678

RESULT 29
US-08-478-373-96
; Sequence 96, Application US/08478373.
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,373
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-478-373-96

Query Match 4.6%; Score 172; DB 2; Length 915;

Best Local Similarity 21.4%; Pred. No. 6.9e-07;

Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

Qy 3 QPMSVFRIN---MTAATVLAALSSVFA--AQTDLETVHIKQORSYNIAIVTEKNGDYSS 57
Db 2 QQQHLFRNLILCLSLMTALPAYAENVQAGQAEQKQDITQVKAQKQKTRRDNEVTG--- 57
Qy 58 PAVTVGTPKIPASLREIPOSVSIITNQVKRNVDTFQRLARKTCLRLVLSNDDGRSSVYA 117
Db 58 -----LGLKLVKTADTLKSKEQVLD-----IRDLTRYDPGIAVVEQGRGASSGS 100
Qy 118 -RGYEYSEYN-IDGLPAQOMS-----INGTLP-----NLFAPDR---VEVMRGPSSG 158
Db 101 IRGMDKNRVSLTVDDL-AQIQSYTAQAALGGTRTAGSSGAINIEIYENVKAVEISKGSNS 159
Qy 159 LFDSSGEMGGIVNLVRKRTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVGRVMA 217
Db 160 VEQSGGALAGSV-----AFQTKTADDVIGEGRGWGIQSKTAYSGKNRGLTQSIALA 210
Qy 218 QTVGASPRPAEKNNRHETFYAAADWDINDPTVLGAGLYQOORHLAPY----- 264
Db 211 GRIGGAELLITGRHA-----GEIRAHEAAGRG-VQSFNRLAPVDVDSKYAYFIVEE 262
Qy 265 ---NGLPADANNKLPSPHOFVFGADW-----NKFKNM---SHDV----- 298
Db 263 ECKNG-----GHEKCKANPKKDVVGEDKQTVSTRDYTGPNRFLADPLSYESRSLPRPGP 318
Qy 299 -FADLKHYFGNGGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRMTVPAPLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
Qy 344 GCNTADKACAVG--LGTEIKQKALAFDASYSRPFLGNTANEFVIGADYNNFRSTNEQG 401
Db 378 GLFTSGENNAPVGAEGYGVDFETHKSRYLEVYVYNADKD--TWADYARLS----- 429
Qy 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKGVGYSHTVATENLDEFG-----IYKSTFFH 456
Db 430 ----YDRQIGLDNHFO---QTHCSAD---GSDKCYRPSADKPPSYKSDRVIYGES--- 476

```

Qy 457 PADGLSLIGGRLGHYKISSGEGKTLHKASKYF-----TGAGAVYDINDNN 504
Db 477 -----HKLLQAAFKKSPDTAKIRHNLVNLGYDRFGSNLRHOD 514
Qy 505 SLVLSLSQLYTPQTNLDADGKLLKP--ROGNOFEVYKGSYMDRLNARVSFYRMKDXXA 562
Db 515 YYYQSANRAVSLKTPPQNNGKKTSPNGREKNPWYSI-----GRNVVTRQICLFGNNT 568
Qy 563 AAPLNPN--NKTRVAALGRVMEGVETEISGAVTPKM-QIHAGYSY---LHS---QIK 612
Db 569 YTDCTPRSINGKSYAAVDNVLG-----RMADVGLAYDRSTHSDGGSYS 617
Qy 613 TASN---SRDDGIFLLMPKHSANLWTTTQVTPPELLTIGGCVNAMSGITSSAGMHAGGYATF 669
Db 618 TGHRTLSNNAGIVL---KPADWLDTYR-----TSTGFRLPBSFA-- 654
Qy 670 DAMAAVFTPKLK-LOINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSPNK 678

RESULT 30
US-08-474-671-96
; Sequence 96, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Loomore, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murrin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:V9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-474-671-96

Query Match          4.6%; Score 172; DB 3; Length 915;
Best Local Similarity 21.4%; Pred. No. 6.9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

Qy 3 QPMVSVERIN--MTAATVAALSSVFA--AQTADLETVIHGKSSYNALVTEKNDDYSS 57
Db 2 QOQHLEFRNLILCLSLMTLPAVAENVQAGQAEKQDLTIQVAKKQKTRDNEVVG---- 57
Qy 58 FAVTVGTKIPASLREIPQSVSIITNQVDRNVDPFDQARKTPGLRLVLSNDGSSVYA 117
Db 58 -----LGKLVKATDILSKQVLD-----IRDLTRYDPGIAVVEQGRGASSGYS 100
Qy 118 -RGYEISEYN--IDGLPAMQMS-----INGTLP-----NLFAFDR---VEVWRGPGG 158
Db 101 IRGMDKNRVSLEYVDGL-AQIQSYTAQALGCTRTGSSGAINIEIYENKAVEISKGSNS 159
Qy 159 LFDSSGEMGIYNLVKRPETKAFQGHAAAG-FGTHKQYKAADVSGSLNSDGSVRGRVVA 217
Db 160 VEOGSGALAGSV-----AFQTKTADVDVIGEGRWGQIGKTAVSGNRGLTQSIALLA 210
Qy 218 QTVGASPRPAEKNNRHETTYAADWDINDPVLGAGYLYQQRHLAPY----- 264
Db 211 GRIGAEALLIRGRLA-----GEIRAEHAARG-VQSFNRLAVDDGSKYAFIVEE 262
Qy 265 ---NGLPADANNKLPSELPHVYVYGDW-----NKRKNM--SHDV----- 298
Db 263 ECKNG-----GHEKCKANPKKDYVEDKQTVSTRTDYTGPNRLAPLSYESSRWLFRPGF 318
Qy 299 -FADLKHFGNGGKGVKGRYSDBDA-----DSNVPAGSKLQMKTPAGR---P 343
Db 319 RRENKHHYIG-GILERTQOTFOTRDNTVPAFLTKAVFDNMQOAGSLRNGKTYAGHKKG 377
Qy 344 GCNTADDKACAVG--LGTEIKOKALAFDASYSRPFRLGNTANEFVIGADYNFRSTNEOG 401
Db 378 GLTSGENNAPYGAEBYGVTFDETHTKSRGLGEVYYTNADKD--TWADYARLS----- 429
Qy 402 RTTLVARGLAL-NEFRSLIPQVDLIANNKGVYRGSHVTATENLDEFG---LYGKSTPH 456
Db 430 ---YDRQIGILDNHFO---QTHCSAD--GSDKYCRPADXPFSYKSDRYLYGSS--- 476
Qy 457 PADGLSLIGGRLGHYKISSGEGKTLHKASKYF-----TGAGAVYDINDNN 504
Db 477 -----HKLLQAAFKKSPDTAKIRHNLVNLGYDRFGSNLRHOD 514
Qy 505 SLVLSLSQLYTPQTNLDADGKLLKP--ROGNOFEVYKGSYMDRLNARVSFYRMKDXXA 562
Db 515 YYYQSANRAVSLKTPPQNNGKKTSPNGREKNPWYSI-----GRNVVTRQICLFGNNT 568
Qy 563 AAPLNPN--NKTRVAALGRVMEGVETEISGAVTPKM-QIHAGYSY---LHS---QIK 612
Db 569 YTDCTPRSINGKSYAAVDNVLG-----RMADVGLAYDRSTHSDGGSYS 617
Qy 613 TASN---SRDDGIFLLMPKHSANLWTTTQVTPPELLTIGGCVNAMSGITSSAGMHAGGYATF 669
Db 618 TGHRTLSNNAGIVL---KPADWLDTYR-----TSTGFRLPBSFA-- 654
Qy 670 DAMAAVFTPKLK-LOINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSPNK 678

RESULT 31
US-08-483-577A-96
; Sequence 96, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Hartness, Robin

```

APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murchin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESS: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,577A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-511
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-577A-96

Query Match 4.6%; Score 172; DB 3; Length 915;
Best Local Similarity 21.4%; Pred. No. 6.9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;
QY 3 QFMSVPRIN--MTAATVLAALSSVFA--AQTADLETVHIKQSRYSNAIVTEKNGDYSS 57
DB 2 QOHLFRILNLCISLMTALPAENVAQQAQEKQLDTIQVRAKKQKTRDNEVTG---- 57
QY 58 FAVTVGTGKIPASLREIPQSVSIITNQVDRNVDTFDQLARKTPGLRLVLSNDDGRSSVYA 117
DB 58 -----LGLVKVTADTILSKQVLD-----IRDLTRYDPGLAVVEQGRGASSGYS 100
QY 118 -RGYEYSEVN--IDGLPAQMS-----INGTLP-----NLFAPDR---VEVMRGPSG 158
DB 101 IRGMDKNRVSLTVDEL-AQIQSYTAQAALGGTRTAGSSGAINEIYENYKAVEISKSGNS 159
QY 159 LFDSSGEMGVNLVRKRTPTKAFQGHAAAG-FGTHKQYKABADVSGSLNSDGSVGRVYMA 217
DB 160 VEQGSALAGSV-----AFQTKTADTVIGEGRWQIGQSKTAVSGKNRGLTQSIALA 210
QY 218 QTVGASPRPAEKNNRHETFYAAADMDINPTDVLGAGLYQQRHLPAY----- 264
DB 211 GRIGGAELLIRTGRHA-----GEIRAHAAGRG-VQSFNRLAPVDGSKYAVFIVEE 262
QY 265 ---NGLPADANNKPSLPQHVFGADW-----NKFPMN--SHDV----- 298

Db 263 ECKNG-----GHEKCKAMPKDVGVGDKRQTVSTRTDYTPGNRFLADPLSYESRSLWFRGPF 318
QY 299 -FADLKHYFGNGGKVGMYRSDRA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG--GILERTQOTFTDTRMTVPAFLTKAVFDANQKQAGSLRGNCKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPRRLGNTANEFVIGADYNNRFRSTNEQG 401
Db 378 GLFTSGENNAFVGAEGYGTGVFDETHTKSRYGLEVVVTNADKD--TWADYARLS----- 429
QY 402 RTTLYARGGLAL--NEPERSIPOVDLIANARKGVRSYSHVTATENLDEFG-----LYKSTFH 456
Db 430 -----YDROGIGLDNHQ---QTHCSAD---GSDKYCRPSADKPFYSYKSDRVIIYES--- 476
QY 457 PADGLSLITGGGRLGHYKIESGEGKTLHKASKTKF-----TCYAGAVYDLNDNN 504
Db 477 -----HKLQAAPKSPDTAKIRHNLSVNLGYDRFGSNLRHQD 514
QY 505 SLYLSLSOLYTPQNLNLDADGKLLKP--RQGNQFVGVYKGSYMDRLNARVSFYRMKQKNA 562
Db 515 YYYQSANRAYSLKTPPQNNKKTSPNGREKNPYWVSI-----GRGNVVTTRQICLFGNNT 568
QY 563 AAPLNP--NKKTRYAALGKRVMEGVETEISGAVTPKW-QIHAGVSY-----LHS---QIK 612
Db 569 YDCTPRISNGKSYAAVRDNRVLG-----RWADVAGLRVYDRSTHSDGGSVS 617
QY 613 TASN---SRDDGIFLLMPKHSANLWTTVQVTPPELTIGGVNAMSGITSSAGMHAGGYATF 669
Db 618 TGTHRTLSWNAIGVL---KPADWLDLTVR-----TSTGFRLPSPA-- 654
QY 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSFNK 678
RESULT 32
US-08-897-438-96
Sequence 96, Application US/08897438
Patent No. 6262016
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murchin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,438
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/483,577
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994

1 PRIOR APPLICATION DATA:
 2 APPLICATION NUMBER: US 08/175,116
 3 FILING DATE: 29-DEC-1993
 4 PRIOR APPLICATION DATA:
 5 APPLICATION NUMBER: US 08/148,968
 6 FILING DATE: 08-NOV-1993
 7 ATTORNEY/AGENT INFORMATION:
 8 NAME: Stewart, Michael I
 9 REGISTRATION NUMBER: 24,973
 10 REFERENCE/DOCKET NUMBER: 1038-720
 11 TELECOMMUNICATION INFORMATION:
 12 TELEPHONE: (416) 595-1155
 13 TELEFAX: (416) 595-1163
 14 INFORMATION FOR SEQ ID NO: 96:
 15 SEQUENCE CHARACTERISTICS:
 16 LENGTH: 915 amino acids
 17 TYPE: amino acid
 18 STRANDEDNESS: single
 19 TOPOLOGY: linear
 20
 21 US-08-897-438-96

Query Match	4.6%	Score 172;	DB 4;	Length 915;
Best Local Similarity	21.4%	Pred. No. 6.9e-07;		
Matches 172;	Conservative 100;	Mismatches 291;	Indels 242;	Gaps 43

[illegible][illegible]

RESULT 33
US-08-637

Sequence 96, Application US/08637654
Patent No. 6358727
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Harkness, Robin E
APPLICANT: Schryvers, Anthony B
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Mordin, Andrew D
APPLICANT: Klein, Michael H
TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENE
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS: 147

[illegible]

QY 159 LFDSSGEMGGIVNLVRKPTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVRGVRMA 217
Db 160 VEOGSGALAGSV-----AFQTKTADDDVIGEGRWGQIOGKTAAYSGKNRGLTQSIALA 210
QY 218 QTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAGLYQOORHLAPY-----264
Db 211 GRIGGAELLIRTRHA-----GEIRAHEAARG-VQSFNRLAPVDDGSKYAYFIVEE 262
QY 265 ---NGLPADANNKLPSPQHVFGADW-----NKFKNM--SHDV-----298
Db 263 ECKNG---GHEKCKANPKKDVGVGDKRQTVTRDYTGPNRFLADPLSVESRSLWPRPGF 318
QY 299 -PADLKHYFGNGGYGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRDVTMPAFLTAKVDFANQKQAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPRFNGTANEFVIGADYNRFRSTNEQG 401
Db 378 GLFTSGENNAAPVGAEGYGVFDETHTKSRYLEYVYTNADK--TWADYARLS-----429
QY 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKVGRGYSHTVATENLDEFG----IYKSTFFH 456
Db 430 ----YDRQGIGLDNHFO---QTHCSAD---GSDKYCRPSADKPFYSKSDRVIYGES---476
QY 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKTKF-----TGAGAVYDLNNDN 504
Db 477 -----HKLQAAPFKKSFDTAKIRHNLSVNLGYDRFGSLRHQD 514
QY 505 SLVLSLSQLYTQTNLDADKLLKP--RCNQPFYGVYKGSYMDRLNARVSPYRMDKNA 562
Db 515 YYYQSANRAYSLKTPQNNQKTSNGREKNPYWVS-----GRGNVTRQICLFGNNT 568
QY 563 AAPLPN--NKTRVAALGKRVMEGVETEISGAVPKM-QIHAGYSY-----LHS---QIK 612
Db 569 YDCTPRSRNGSKSYAAVRDNRVLG-----RWADVAGRLYDRYRSTHSDGSGVS 617
QY 613 TASN---SRDDGIFLLMPKHSANLWTVYQVTPELTIGGVNAMSITTSAGMHAGGYATF 669
Db 618 TGTHRTLSNAGIVL---KPADWLDLTVR-----TSTGRLPSFA--654
QY 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSFNK 678

RESULT 34
US-08-649-518-96
; Sequence 96, Application US/08649518
; Patent No. 6361779
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Fele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/649, 518
;; FILING DATE: 17-MAY-1996
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/337,483
;; FILING DATE: 08-NOV-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/483,577
;; FILING DATE: 07-JUN-1995
;; APPLICATION DATA:
;; APPLICATION NUMBER: US 08/175,116
;; FILING DATE: 29-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/148,968
;; FILING DATE: 08-NOV-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stewart, Michael I
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-608
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1155
;; TELEFAX: (416) 595-1163
;; INFORMATION FOR SEQ ID NO: 96:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 915 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-649-518-96

Query Match 4.6%; Score 172; DB 4; Length 915;
Best Local Similarity 21.4%; Pred. No. 6.9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

QY 3 QFMSVPRIN---MTAATVLAALSSVFA--AQTADLTGVHIKQORSYNAIVTEKNGDYSS 57
Db 2 QOQHLFRNLILCLSLMTALPAYAENVQAGQAEKOLDTIQVKKAKKQTRRDNEVTG----57
QY 58 PAVTVGTGKIPASLREIPQSVSIITNQVKDRNVDTFQQLARKTPGLRVLNSDDGRSSVYA 117
Db 58 -----LGKLVKTADTSLKEQVLD-----IRDLTRYDGIADVVEGRCASSGYS 100
QY 118 -RGYSEYN--IDGLPAQMS-----INGTLP-----NLFAPDR---VEVMRGPFG 158
Db 101 IRGMDKNRVSLTVDDL-AQIQSYTAQAALGGTRTAGSGAINETEYENVKAVEISKGSNS 159
QY 159 LFDSSGEMGGIVNLVRKPTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVRGVRMA 217
Db 160 VEOGSGALAGSV-----AFQTKTADDDVIGEGRWGQIOGKTAAYSGKNRGLTQSIALA 210
QY 218 QTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAGLYQOORHLAPY-----264
Db 211 GRIGGAELLIRTRHA-----GEIRAHEAARG-VQSFNRLAPVDDGSKYAYFIVEE 262
QY 265 ---NGLPADANNKLPSPQHVFGADW-----NKFKNM--SHDV-----298
Db 263 ECKNG---GHEKCKANPKKDVGVGDKRQTVTRDYTGPNRFLADPLSVESRSLWPRPGF 318
QY 299 -PADLKHYFGNGGYGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRDVTMPAFLTAKVDFANQKQAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPRFNGTANEFVIGADYNRFRSTNEQG 401
Db 378 GLFTSGENNAAPVGAEGYGVFDETHTKSRYLEYVYTNADK--TWADYARLS-----429
QY 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKVGRGYSHTVATENLDEFG----IYKSTFFH 456
Db 430 ----YDRQGIGLDNHFO---QTHCSAD---GSDKYCRPSADKPFYSKSDRVIYGES---476
QY 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKTKF-----TGAGAVYDLNNDN 504

Db 477 -----HKLQAAFKKSFDTAKTRHNLISVNLGIDRFGSNI.RHOD 514
Qy 505 SLVLSLQVTPOTNLDADGKLLK--ROGNOPEVGYKGYMDRLNARVSFYRMKDKN 562
Db 515 YYYQSANRAYSLLKPPQNNKKTSPNGREKNPFWSI-----GRNVVTRQCLGNNT 568
Qy 563 AAPLNPN--NKTRVYALGKRVMEGVETEISGAVTPKW-QIHAGYSY---LHS---QIK 612
Db 569 YTDCTPRISNGKSYAAVDNVLG-----RMADVAGLRDYRSTHSDGSVS 617
Qy 613 TASN---SPDDGIFILMPKHSANLMTTYQVTPPELLITGGGNMNSGITSSAGMAGYATP 669
Db 618 TGHRTLSNMAGIVL--KPADWLDTLYR-----TSTGFRLLPSFA-- 654
Qy 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGMRSGDKIKAVKIDPEKSPNK 678

RESULT 35
US-08-363-124A-4
; Sequence 4, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 909 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-363-124A-4

Query Match 4.5%; Score 171; DB 2; length 909;
Best Local Similarity 18.2%; Pred. No. 8, 4e-07;
Matches 180; Conservative 128; Mismatches 327; Indels 356; Gaps 49;

Qy 3 QFMSVFRINMTATVLAALSSVFA-----AQADLETVHIKGRSYNAIVTERKNGY 55
Db 2 QOQHFLRLNILCLISLTAL--PYAENVOAQEQKQLDITIYKAKKQKTRRDNEVTG-- 57
Qy 56 SSFATVYVGRKIPASLEIRIQSVSIITNOQYKBRNDVTPQQLAKRTGLAVLSDDRSV 115
Db 58 -----LGLVKSSTPTLSKEQVLN-----IRDLTRYDGIIVVEGGRASSG 98
Qy 116 YA-RGYEYSEVN--IDGLPAQOMOS-----INGTLP-----NLPAFR--VEYWRG 156
Db 99 YSIRGMDKRRVSLTVDGV-SQIOSTYAQAALGSTRAGSSGALNEIYENAVAEISKGS 157
Qy 157 SGLFDSSGEMGGIVN----- 171
Db 158 NSSEYNGNLAGSVAFQRTAADIIIEGKOWIGISKTAVSGKHDLTQSLALAGSGAE 217
Qy 172 --LVKRPYKAFQGHAAAGFGTH-----KQYAEADV-----G 203
Db 218 ALLIYTKRRGRREIHAHKDAGKGVOSFNRLVLDEDDKKEGSGSDISLCEBECHNGYAACN 277
Qy 204 SLNSDGSV--RGRVVAQTVGASPR---PAE-----KNRH----- 233
Db 278 KLKEDASVDERKTVSTDYTGSNRLANPLEYGSQWLFRPGMHLDNRHYVGAIVLETRQ 337
Qy 234 ETP-----YAADWDINP-----DTVLGAGY---LY 256
Db 338 QTFDDTDMVPAVFTSE-DYVGSGLKGLKRGSDNKAERLFPVQGESESTIGIGYGVFPY 396
Qy 257 QQRHLAPYNGLPADANNKLPSPQHYFVGADNKKFMNSHDVFDLKHYPGNGYGRVGM 316
Db 397 DERHTKRRYGV-----EYVYNAD-----KDTWAD---YARLSYRQGI 432
Qy 317 RYSDRPADSNVAFAGSKLGKMTKTPAGRGCTADDKACAVGLGEIKOKALAPASSRPE 376
Db 433 DLDNRLQOTHCSDGSDKCRPDGKPYFSYSDRNI-----YESRNLPAVVKKAF 485
Qy 377 RLGNTEANEFVIGADYNFRSTNEQGRITLYARGGLANFRSIPQVLDIANNR---KGV 432
Db 486 DTAKIRHNLSINLGYDRFS--QLSHSDVYLL-----NAVQADLTTPKKPPEPNQS 535
Qy 433 RGYSHIVA---TENIDFGIYKSTFFHPADGLSLIGG-----RLGHYKIESGEG 479
Db 536 KNPYFVYSIGKTTVNTSPPIPGFNNTYTCTPRNIGANGYVAAVDNVLGRW-ADVGA 594
Qy 480 -----KTLKASKTKFTG-----YAGAV-----YDLNDNNSLYL---SLSQLTPTQ 518
Db 595 IRDYDSTHSSEDSKSVSTGTHRNLSMNAGVLLKPTWMDLTYRSTGFRLLPSFEMYGWRA 654
Qy 519 NLDADGKLLKPRQGNQFEVG--YKG-----SYMDRLNARVSF-YRMKDKN-----AA 563
Db 655 GBSLKTLDLKPEKSPYREAGIYFPGDFGMLASVFNNAVRDLAFQYETRTQNGQTSASG 714
Qy 564 APINPNKKTRYAA---LGRVMEGVETEISGAVTPKW---QIHAGYSYLSHQIKTA 614
Db 715 DPGYRNGPKCTVAGINILGKIDMHGV-----WGLPDGLYSTLAVNRIRYKDA 763
Qy 615 SNRSDGIFL-----LMPK-----HSANLM-----TTYO-----VTBELTI 645
Db 764 -DIRADRTVTSYLFDAVOPSRVYGLGYDHPDGMGINTMFTFYSYAKSVDELISQALL 822
Qy 646 GGGVNMASGITSSAGMHAGGYATFDAMAAYRFTPKLKLOINADNIFNRHY----- 695
Db 823 NGNANA---KKAASRRTRPMVYTVDSGYNNIKKHLLTLAGAGVNNLNTYVYVWENVRQTA 878
Qy 696 -----YARVGGANTFNIPGSERTTANLRY 720
Db 879 GGAVNQHKVGVYNRVAAADGRNVTFSLEMKF 909

RESULT 36
US-08-817-707-8
; Sequence 8, Application US/08817707
; Patent No. 6277382

```
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Hefron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 790 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-707-8

Query Match          4.5%; Score 169.5; DB 4; Length 790;
Best Local Similarity 18.5%; Pred No. 9,2e-07;
Matches 153; Conservative 102; Mismatches 336; Indels 235; Gaps 31;

Qy 18 LAALSSSVF-----AAQTADLETVHIKQRSYNAIVTERKNGDYSSFAVTVGKIPASLRE 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9  IAAVGSIFGNPVLAADEAATTPVKAEIKVVRKQQLNAPATVERVNLGRIQDEMIRD 68

Qy 73 IPQSVSIITNQVKRNVDTPQLAKTFLGLRVLSNDDG-----RSVYARGYEYSEY 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 NKDLVRYSTDVGLSDSGRHQKGFVARGVGVSGVSDGVSLPDSSENSLYAR---YGNF 125

Qy 126 NIDGPAQWQSQINGTLPLNLFADRVFVMEVSGSLFDSSGEMGGINLVLRKRTKAPQGH- 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 NSSRLSIDPELVR-----NIEIAKGADSFNTGSCALGGVNY-----QTLQGH 169

Qy 185 -----AAAGFTGHKQYKAEADVSGSLNS 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 LLLDDRQFGVMKNGVSSRNREWTNLTGFGVSNDRVDAALLYSQRRGHETESAGERGPV 229

Qy 208 DGSVGRVMAQTVGASPRPAEKNNRHETPYAAADWDINDPTVLGAGLYQQRH----- 260
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 EGAGSGAIIRSGSRGTPDPSPK--HKYHNPFLGKIAYQINDKHRIQGFSGNQOQHNTYIBES 287

Qy 261 --LAPVNGLPADANKPLPSLPQHVFGVADWKNFKMNSHVDVFDLKH-----FGNGG-- 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 YNLTSASSWEADVNR-----RRANLFTVPTDPSNWLSSLKADFDYQYTKVAAVNKNKSF 343

Qy 311 ---YGVKVMRYSDRSDASNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTETIKQKALA 367
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 PTDTSTWTNRYNQKLENIY-----NRSMD-----TRFKRFTLR 377

Qy 368 FDASYRPPRLNGLTNEFVIGADYNRSTNEQGRITLYARGGLALNEFRSIPQVDLIAN 427
```

```
Db 378 MD---SQPLQLG-----GHRLSLKTFSRRREFENLNRDDYYFS 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 428 AR--KGVRGYSGHTVATENLDEFGYIGKSTFHPADGLSLIGGRLGHYKIESGE-GKTLHK 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 ERVSRTTSSIQHPVKTTN---YGFSLSDQIQWNVFSSRADIRYDHTKMTQELNADCHA 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 485 ASKTK-----FTGYAGAVYDLN-----DNNSLYL--SLSQLY-----TPQ 517
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 CDKTPPAANTYKWSGFGVGLAAQLNQHWHVGYDITSGYRVFVNAASEVFTYNHSGSNMLPN 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 518 TNLDA-----GKLLK-PRQNOFQVGYKGSYMDRLNARVS-----PYR 556
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 PNLKAERSTTHTLSLQGRSEKGTLDANDYQNNYR-NFLUSEEQNLTTSDGVCCTQNNYYIG 589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 557 MKDKNAAPLNPNNKTRYAA--LGKRVMEGVE-----TEISGAVTPKQIHAAGSYL 607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 MCS-----NPYSEKPEWQMNIDKARIRGLELTGRLNVTKVASFVPEGKLFGLSLGYA 642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 608 HSQ-----IKTASNSRDDGIFLLMPKXHSANLWTTVYQVTPELTIGGGV 649
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 643 KSLSGDNSLLSTQPPKVIAGIDYESPEKMGVFSRLTYLGAKKAKDAQYTVYENKGRGT 702
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 650 NAMSGITSSAGMHAGGYATFDAMAAYRFTPKLKLQINADNIFNRHY 695
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 703 PLQKKVKDYPLWLNKSA--VFDWYGFYKLAKNLTLRAGVYVNVFNRY 747
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 37
US-08-613-009A-16
; Sequence 16, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-613-009A-16
```


Query Match 4.5%; Score 169.5; DB 3; Length 908;
 Best Local Similarity 18.0%; Pred. No. 1.2e-06;
 Matches 177; Conservative 125; Mismatches 341; Indels 339; Gaps 46;

```

QY 3 OPMVSFRIMTAATVLAALSSSVFA-----AQTADLETVHIKGORSYNAIVTEKNGDY 55
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 2 QOOHLFRILNLCISLMTAL--PYVAENVOAEOAQOEKQDITIOVAKKOKTRDNEVTG-- 57
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 56 SSFAVTVGTXIPASLREIPQSVSIITNOQVKNRVNDFDQIARKTPGLRYLSNDGRSSV 115
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 58 -----LQKLVKSSDPTLSKEQVLA-----IDLTRYDYGIAVVEQGRASSG 98
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 116 VA-RGEYSEYN--IDGLPAQOMOS-----INGTLP-----NLFAFPR---VEVMRGP 156
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 99 YSIRGMDKNRVSILTYDGV--SOIOSYTAQAALGSTRAGSSGAINETIENVAKAVEISKGS 157
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 157 SGLFDSSGEMGGIVN----- 171
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 158 NSSEYONGALASVAFQOTTAADIIGEGKQWGIQSKTAVSGKHALTQSLALAGSGGAE 217
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 172 ---LVKRPPTKAFQGHAAAGFGTH-----KQYK-----AEADVSG 203
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 218 ALLIYTKRGRREIHAHKDAGKGVQSFNRLVLDEKKEGSGYRYFIVEBECHNGYAACKN 277
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 204 SLNSDGSV---RGRVMAQTVGASPR---PAE-----KNNRH----- 233
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 278 KIKEDASVDERKTVSTQDYTGNSRLANPLEYGSQSWLFRPGMHLDNHRHYGAVLERTQ 337
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 234 ETF-----YAADWDINP-----DTVLGAGY-----LY 256
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 338 QTFTRDMTPVPAFTTSE--YVVPQSLKGLGKYSQDNKAERLFOGEGSTLQGIQYGTGVFY 396
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 257 QORHLAPYNGLPADANNKLPSPLOHVFVAGADWKKFKONSHDVPAHLKHYFGNGGYKVGKA 316
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 397 DRRHTKNRKGV-----EYVYHNAD-----KOTWAD-----YARLSYRQGI 432
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 317 RYSDRDADSNVAFAGSKLQMKTPAGRPGCNTADDKACAVGLTEIKOKALPADASYRPF 376
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 433 DLNDRLOQTHCHSDGDKCRPDGKNKPYSPYKSDRM1-----YESRNLFOAVFKKAF 485
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 377 RLGNTANERVLGADVNRFRSTNQGRTILYARGLALNFRSTPOVDILANARKGYRGIS 436
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 486 DTAKTRHNLSINLGYDREKS--QLSHSDYLLQNAVOAYDLITPPKPPFGSGKDNRYRVS 543
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 437 HTVATENLDEFGIYGKSTPHPADGLSLIGSG-----RLGHYKIESGEG-----KTL 482
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 544 ICKTYVNTSPICRFGNNYTTDCTPRNIGNGGYAAVQDNVRLGRW--ADVAGAGIRYDRST 602
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 483 HKASKTKFTG-----YAGAV-----YDLNDNNSLYL--SLSQLYTPQTNLDADGKL 526
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 603 HSEDKSVSTGTRHNLNLSMAGVVLKPFTHMDLTFRASTGFRLLPSFALMYGWRAGESLKTLD 662
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 527 LKPRQONQEVG--YKG-----SYMDRLNARVSF--YRMADKN-----AAAPLPNNK 571
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 663 LKPEKSFNREAGIVFGDGNLEASVFNNAYRDLAFAGYETRTQONQTSAGDPGRNRQ 722
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 572 KTRVAA--LGRVMEGVETEISGAVTPKX-----OIHAGSYLSHQITVANSRSDGJF 623
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 723 NARIAGINILGKIDMHGV-----WGGIPLDGLYSLTAVNRLKAVDA--DIRADRIF 770
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 624 L-----LMPK-----HSANLM-----TTYQ-----VTPELLTGGGVANMSG 654
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 771 VTSYLFDAVQPRRYVLGLGYDHPDGIMGTMTFTYSKASVDELLASQALLNQNANA----- 827
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 655 ITSSAGMHAGVATPDMAAAYRFTPLKQLINADNIFNNHY-----YAR 698
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 828 -KKAASRRTRPMTVDVSGVYNIKKHLLTLAGVYLLNRYVTWENVROTAGAVNQHKN 886
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 699 VEGANTFNIPGSERTWTANLRY 720
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 887 VGVYNRYPAAFGNRYTFSLEKPF 908
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 38

US-08-778-570B-22
 ; Sequence 22, Application US/08778570B
 ; Patent No. 6437096

GENERAL INFORMATION:

APPLICANT: Myers, Lisa E
 APPLICANT: Schryvers, Anthony B
 APPLICANT: Harkness, Robin E
 APPLICANT: Loosmore, Sheena M.
 APPLICANT: Du, Run-Pan
 APPLICANT: Yang, Yan-Ping
 APPLICANT: Klein, Michel H
 TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sim & McBurney
 STREET: 6th floor, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/778,570B
 FILING DATE: 03-JAN-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I
 REGISTRATION NUMBER: 24973
 REFERENCE/DOCKET NUMBER: 1038-664
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 908 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-08-778-570B-22

Query Match 4.5%; Score 169.5; DB 4; Length 908;
 Best Local Similarity 18.0%; Pred. No. 1.2e-06;
 Matches 177; Conservative 125; Mismatches 341; Indels 339; Gaps 46;

```

QY 3 OPMVSFRIMTAATVLAALSSSVFA-----AQTADLETVHIKGORSYNAIVTEKNGDY 55
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 2 QOOHLFRILNLCISLMTAL--PYVAENVOAEOAQOEKQDITIOVAKKOKTRDNEVTG-- 57
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 56 SSFAVTVGTXIPASLREIPQSVSIITNOQVKNRVNDFDQIARKTPGLRYLSNDGRSSV 115
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 58 -----LQKLVKSSDPTLSKEQVLA-----IDLTRYDYGIAVVEQGRASSG 98
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 116 VA-RGEYSEYN--IDGLPAQOMOS-----INGTLP-----NLFAFPR---VEVMRGP 156
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 99 YSIRGMDKNRVSILTYDGV--SOIOSYTAQAALGSTRAGSSGAINETIENVAKAVEISKGS 157
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 157 SGLFDSSGEMGGIVN----- 171
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 158 NSSEYONGALASVAFQOTTAADIIGEGKQWGIQSKTAVSGKHALTQSLALAGSGGAE 217
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 172 ---LVKRPPTKAFQGHAAAGFGTH-----KQYK-----AEADVSG 203
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 218 ALLIYTKRGRREIHAHKDAGKGVQSFNRLVLDEKKEGSGYRYFIVEBECHNGYAACKN 277
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 204 SLNSDGSV---RGRVMAQTVGASPR---PAE-----KNNRH----- 233
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 278 KIKEDASVDERKTVSTQDYTGNSRLANPLEYGSQSWLFRPGMHLDNHRHYGAVLERTQ 337
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

QY 234 ETP-----YAADWDINP-----DTVLGAGY-----LY 256
Db 338 QTFTDRMTVPAYFTSE--DYVPGSLKGLGKYGDNKAERLFGQEGSTLQGIYGTGVFY 396
QY 257 QORHLAPYNGLPADANNKLPSPHOFVGVADWNKFMNSHDVFADLKHVFGNGGKVGVM 316
Db 397 DERHTKNRYGV-----EYVHNAD-----KOTWAD-----YARLSYDRQGI 432
QY 317 RYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 376
Db 433 DLNRLQOQTHCSHDGSKNCRPDGNKPYFYKSDRMI-----YEESRNLFOAVFKKAF 485
QY 377 RLGNANEFVIGADYNRFRSTNEQGRITLYARGGLALNEFRSIPQVDLIANARKVGRVYS 436
Db 486 DTAKIRHNLINLGYDRFKS--QLSHSDYLLQNAVQAVDLYTPPKPPFNGSKNPNYRVS 543
QY 437 HTVATENLDEFGYKSTFHPADGLSLIGG-----RLGHYKIESGEG-----KTL 482
Db 544 IGKTTVNTSPICRFGNNTYDCTPRNIGNGYAAVQDNVRLGRW-ADVGAIRYDYRST 602
QY 483 HKASKTKFTG-----YAGAV-----YDLNDNNSLYL---SLSQLYTPOTNLDAQGL 526
Db 603 HSEDKSVSTGTHRNLNWNAGVVLKPTWMDLYRSTGFRLPSPFAEMYGWRAGESLKTLD 662
QY 527 LKPRQNGQEVG--YKG-----SYMDDLNRARVSF-YRMKDN-----AAAPLNPNKK 571
Db 663 LKPEKSFNREAGIVFKGDFGNLEASYFNNAVRDLIAFGYETRTONGOTSASGDPGRNAQ 722
QY 572 KTRYAA--LGRVMGVEVETISGAVTPKW-----QIHAGYSLHSQIKTASNRDDGIF 623
Db 723 NARIAGINILGKIDHGV-----WGLPDGLSTLAYNIKVKADA--DIRADRTF 770
QY 624 L-----LMPK-----HSANLW-----TTYQ-----VTPELTIGGVNMSG 654
Db 771 VTSYLFDAVQPSRYVLGLYDHPDGIWGTINTWFTYSKAKSVDELLGSQALLNGNANA---827
QY 655 ITSSAGMHAGGYATPDMAAYRFTPKLKLQINADINFRHY-----YAR 698
Db 828 -KKAASRRTRPWPVTDVSGYINIKKHLTLRAGVYNLLNRYVYTWENVRQTAGGAVNOHKN 886
QY 699 VGGANTFNIPGSERTWTANLRY 720
Db 887 VGVNRYAAPGRNYTFSLEMKF 908

RESULT 39
US-09-584-22

; Sequence 22, Application US/09059584

; Patent No. 6440701

; GENERAL INFORMATION:

; APPLICANT: Myers, Lisa E

; APPLICANT: Schryvers, Anthony B

; APPLICANT: Harkness, Robin E

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Du, Run-Pan

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/059,584

; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-059-584-22

Query Match 4.5%; Score 169.5; DB 4; Length 908;
Best Local Similarity 18.0%; Pred. No. 1.2e-06;
Matches 177; Conservative 125; Mismatches 341; Indels 339; Gaps 46;

QY 3 QMSVFRINMTAATVLAALSSVFA-----AQTADLETVHIKQRSYNAIVTEKNGDY 55
Db 2 QOQHLFRNLCLSLMTAL--PVYAENVQAEQAEKQLDTTQVKAKKQKTRDRNEVTG-- 57
QY 56 SSFATVGTCTKIPASIREIPQSVSLITNQVKDRNVDPDQLARKTPGLRVLNSDDGRSSV 115
Db 58 -----LGLVKSSDLSKEQVLN-----IRDUTRDPGIAVVEQGRGASSG 98
QY 116 YA-RGEYSEYN--IDGLPAQMQS-----INGTLP-----NLFAFDR---VEVMRGP 156
Db 99 YSRGMDKNRYSLVTDGV-SQIQSYTAQAALGGTRTAGSSCAINEIEVENYKAVEISKGS 157
QY 157 SGLFDSGEMGGIVN-----SGLFDSGEMGGIVN-----SGLFDSGEMGGIVN 171
Db 158 NSSEYGNALAGSAVAFQTKTAADIIGEGKQWGIQSTAYSGKDHALQSLALAGRSGAE 217
QY 172 ---LVRKPTKAFQGHAAAGFGTH-----KQYK-----AEADVSG 203
Db 218 ALLIYTKRGRREIHAHKDAGKGVQSFNRLVLEDDKGGSGQRYRVIIEEBCHEGNYAACN 277
QY 204 SLNSDGSV---RGRVMAQTVGASPR---PAE-----KNNRH----- 233
Db 278 KLKEDASVKDERKTVSTQDYTGSNRLLANPLEYGSQSQWLPFGWHLDNRHYVGVALERTQ 337
QY 234 ETP-----YAADWDINP-----DTVLGAGY-----LY 256
Db 338 QTFTDRMTVPAYFTSE--DYVPGSLKGLGKYGDNKAERLFGQEGSTLQGIYGTGVFY 396
QY 257 QORHLAPYNGLPADANNKLPSPHOFVGVADWNKFMNSHDVFADLKHVFGNGGKVGVM 316
Db 397 DERHTKNRYGV-----EYVHNAD-----KOTWAD-----YARLSYDRQGI 432
QY 317 RYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 376
Db 433 DLNRLQOQTHCSHDGSKNCRPDGNKPYFYKSDRMI-----YEESRNLFOAVFKKAF 485
QY 377 RLGNANEFVIGADYNRFRSTNEQGRITLYARGGLALNEFRSIPQVDLIANARKVGRVYS 436
Db 486 DTAKIRHNLINLGYDRFKS--QLSHSDYLLQNAVQAVDLYTPPKPPFNGSKNPNYRVS 543
QY 437 HTVATENLDEFGYKSTFHPADGLSLIGG-----RLGHYKIESGEG-----KTL 482
Db 544 IGKTTVNTSPICRFGNNTYDCTPRNIGNGYAAVQDNVRLGRW-ADVGAIRYDYRST 602
QY 483 HKASKTKFTG-----YAGAV-----YDLNDNNSLYL---SLSQLYTPOTNLDAQGL 526
Db 603 HSEDKSVSTGTHRNLNWNAGVVLKPTWMDLYRSTGFRLPSPFAEMYGWRAGESLKTLD 662

QY 527 LKPRQNOFEVG--YKG-----SYMDRLNARVSF-YRMKDN-----AAPLIPNKK 571
 Db 663 LKPEKFNREAGIVFGKDFGNELEASYNNAAYRDLIAFGYETRTQNGOTSAGDPGRMAQ 722
 QY 572 KTRVAA---LGKRVMEGVETETISGAVTPKW-----QIHAGSYLSHQSOKTASRPDGF 623
 Db 723 NARIACINILGKIDMHGV-----WGLPDLGYSTLAINRIKVKDA-DIRADRTF 770
 QY 624 L-----LMPK-----HSANLW-----TTYQ-----VTPELTIGGVNAMSQ 654
 Db 771 VTSYLFDAVQPSRYVLGLGDHPDGLMGINTMFTYSKAKSVDELISQALINGNANA--- 827
 QY 655 ITSSAGMHAGGYATFPDAMAAYRTPYKLGQINADNIFNRHY-----YAR 698
 Db 828 -KKAASRRTRPWWTVTVSGYINIKKHLLTRAGVYNLLNRYVTWENVROTAGAVNQHN 886
 QY 699 VGGANTFNIPGSERTWTANLRY 720
 Db 887 VGVYNRVYAAPGRNRYTFSLEMKF 908
 RESULT 40
 US-08-487-890A-94
 ; Sequence 94, Application US/08487890A
 ; Patent No. 5708149
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Mardin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,890A
 ; FILING DATE: 07-JUN-1993
 ; CLASSIFICATION: 435
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/175,116
 ; FILING DATE: 29-DEC-1993
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/148,968
 ; FILING DATE: 08-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 94:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 908 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 US-08-487-890A-94

Query Match 4.4%; Score 167.5; DB 1; Length 908;
 Best Local Similarity 18.2%; Pred. No. 1.8e-06;
 Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;
 QY 3 QPMSVRIMTATVLAALSSVFA-----AQTDLFTVHIKGRSINATYTEKNGDY 55
 Db 2 QOQHFLRNILICLSMTAL--PYAENVQAEQAEKQDLDTIQKAKOKTRRDNETGTG-- 57
 QY 56 SSPAATVGTKIPASLEIREIQSVSIINQOVRNVDTPOLARKTGLRYLSNDGRSSV 115
 Db 58 -----LGLVKSDDTSLKEQYLN-----INDLIRYDGIATVEGGRASSG 98
 QY 116 YA-RGYEYSEYN--IDGLPAQMS-----INGTLP-----NLPAFDR--VEVMRGP 156
 Db 99 YSIRGMDKNRVSILTVDGV--SQIQSYTAQAALGTRTAGSSGAINIEYENVKAVEISKS 157
 QY 157 SGLPSSGEMGIVN----- 171
 Db 158 NSEYVNGALAGSVAFQYTAADIIGEKQWGIQSKTAYSGKDHALTOSLALAGRSAGAE 217
 QY 172 ---LVKRPPTKAFQGHAAAGFGTH-----KQYK-----AEADVSG 203
 Db 218 ALLITYKRGREIHAHKDAGKGVQSFNRLVDEBKKEGGSQYRYFVEECHNGVYACKA 277
 QY 204 SLNSDGSV--RGRVMAQTVGASPR---PAE-----KNRH----- 233
 Db 278 KLEKEDASVDERKTAVTQDYTGSSNRLANPLEYSGQSWLFRGMHMDNHVYCAVLERTQ 337
 QY 234 EFT-----YAAADWQINP-----DYLGAGY---LY 256
 Db 338 QTFDTRDMTVPAVFTSE--DYVQSLKGLGKYGSDNKAERLFPVQEGESTIQGIGYGTVEY 396
 QY 257 QORHLAPYNGLAPADANNKLPSPLOHYFVGADWNKFRGNSHDVPAFLKHFGNGYGVKM 316
 Db 397 DEHHTNRYGV-----EYVYNAD-----KDTWAD---YARLSYDRQGI 432
 QY 317 RYSDRPADSNYAFAGSKLGKMTPAGRPGCNTADPKACAVGLGTEIKOKALAPDASRPF 376
 Db 433 DLNRLQOYTHGCHSDGDKNCRPDKNKPYSFYKSDRI-----YESNLLFGAVPKKAF 485
 QY 377 RLGNANAEVIGADINRFRSTNEQGRITLYARGGLALNEFRSIPQVDLIANR---KVY 432
 Db 486 DPAKIRHNLINILGYDRFSS--QLSHSDYLLQ-----NAVQAYDLITPKKPPPPNGS 535
 QY 433 RGVSHVA---TENLDERGIYKGSFHPADGLSLIGG-----RLGHYKLESGBG 479
 Db 536 KNPYRVSIGKTTVNTSPICRFGNNTYDCTPRNIGANGYVAAVQDNVRLGRW-ADVGA 594
 QY 480 -----KTLHAKSKTKPTG-----YAGAV-----YDLNUNSLYL---SLQLYTPQT 518
 Db 595 IRYDYRSTHSEKSVSTGTHRLSNMNAAGVLLKPTTMDLTYRASTGFRLLPSFAEMVGMA 654
 QY 519 NLDADGKLKPRQNOFEVG--YKG-----SYMDRLNARVSF-YRMKDN-----AA 563
 Db 655 GSKLTKLDLKPESFNRAGIVFKGDFGNLEASYNNAAYRDLIAFGYETRTQNGOTSAG 714
 QY 564 APLNPNKKTRYAA---LGKRVMEGVETETISGAVTPKW-----QIHAGSYLSHQSOKTAS 615
 Db 715 DPGRVNAQVARLAGINILGKIDMHGV-----WGLPDLGYSTLAINRIKVKDA- 762
 QY 616 NSRDGIFL-----LMPK-----HSANLW-----TTYQ-----VTPELTIG 646
 Db 763 DIRADRTFTSYLFPDAVQPSRYVLGLGYHPGIGINMFTYSKAKSVDELISQALIN 822
 QY 647 GGVNAMSIGITSSAGMHAGGYATFPDAMAAYRTPYKLGQINADNIFNRHY----- 695
 Db 823 GNANA---KKAASRRTRPWWTVTVSGYINIKKHLLTRAGVYNLLNRYVTWENVROTAG 878
 QY 696 -----YARVGGANTFNIPGSERTWTANLRY 720
 Db 879 GAVNQHNQVYNRVYAAPGRNRYTFSLEMKF 908

RESULT 41

US-08-478-435-94
 ; Sequence 94, Application US/08478435
 ; Patent No. 5922323
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/478,435
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/337,483
 ; FILING DATE: 08-NOV-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/175,116
 ; FILING DATE: 29-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/148,968
 ; FILING DATE: 08-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 94:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 908 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-478-435-94

Query Match 4.4%; Score 167.5; DB 2; Length 908;
 Best Local Similarity 18.2%; Pred. No. 1.8e-06;
 Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

Qy 3 QPMSVFRINWTAATVLAALSSVFA-----AQTADLETVHKQORSYNAIVTEKNGDY 55
 Db 2 QOCHLFRNLCLSLMTAL--PVYAEVQAEQKQOLDIQVAKKQKTRRDNEVTG-- 57
 Qy 56 SSFAVTGKIPASLREIPQSVIITNOQKDRNVDTFQDLARKTPGLRLVLSNDDGRSSV 115
 Db 58 -----LKKVKSDDTSLSKEQVLN-----IRDLTRYDPGTAIVQEGRGASG 98
 Qy 116 YA-RGYEYSEYN--IDGLPAQMOS-----INGTLF-----NLFAFDR-----VEVMRGP 156
 Db 99 YSIRGMKDRVSLTVDGV--SQIQSVTAQAALGGTRTAGSSGAINIEYENKVAIVEISKS 157
 Qy 157 SGLFSSGEMGIYN----- 171

Db 158 NSSEYNGALAGSVAFQTKTAADIIGEGKQWGIOSKTAYSKQKHALTQSLALAGSSGAE 217
 Qy 172 ---LVKRKPTKAFQCHAAAGFGTH-----KQYK-----KQYK-----AEDVSG 203
 Db 218 ALLIYTKRGREIHAHKDAGKGVQSPNRLVLDEKKEGSGYRYFIVEECHNGVAAACKN 277
 Qy 204 SLNSDGSV---RGRVMAQTVGASPR---PAE-----KNNRH----- 233
 Db 278 KLKEDASVKDERKTVSTQDYTGSNRLLANPLFEGSQSWLFRPGWHLNDRHYVGVAVLRTQ 337
 Qy 234 ETF-----YAAADWDINP-----DTVLGAGY-----LY 256
 Db 338 QTFDTRMTVPAYFTSE--DYVPGSLKGLGKYGKGNKAEERLFVQEGSGTLOGIGTGVFY 396
 Qy 257 QORHLAPYNGLPADANNKLPSPHQRFVFGADWNKFKMNSHDFADLKHVFGNGGYGVKGM 316
 Db 337 DERHTKNRYGV-----EYVYHNAD-----KDTWAD-----YARLSYDROGI 432
 Qy 317 RYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTETIKKALAFDASYSRPF 376
 Db 433 DLNRLQOQTHCSHDSKNCRCPDGKNKPYFYSKDRMI-----YBSRNLFOAVFKKAF 485
 Qy 377 RLGNANTANEFVIGADYNRFRSTNEOGRITLYARGGLALNEFESI PQVDLIANAR-----KGV 432
 Db 486 DTAKIRHNLNLSINLGYDRFKS--QLSHSDYLYQ-----NAVQAYDLITPKKPPFPNGS 535
 Qy 433 RGYSHTV-----TENLDFGIYKSTFHPADGLSLIGGG-----RLGHYKIESGEG 479
 Db 536 KDNPRVSIKGTVTNTSPICRFGNNTYDCTPRNIGGNYAAVQDNVRLGRW-ADVGAG 594
 Qy 480 -----KTLHKASKTKFTG-----YAGAV-----YDLNDDNSLYL---SLSOLYTPQT 518
 Db 595 IRYDYRSTHSEKSVSTGTHRNLSWNAQVVLKPFMTWMDLYTRASTGPRLPSPFAEMYGWRA 654
 Qy 519 NLDADGKLLKPRQGNQEVG--YKG-----SYMDDLNAARVSP-YRMKDKN-----AA 563
 Db 655 GESLTKLDLKPESFNREAGIVFKGDFGNLEASYFNNAAYRDLIAFGYETRTQNGOTSASG 714
 Qy 564 APLNPNKKTRYAA---LGKRVMEGVETEISGAVTPKW-----QIHAGYSYLHSHQIKTAS 615
 Db 715 DPGYRNAQNARIAGINILGKIDWHGV-----WGLPDLGLYSTLAYNRILKVKDA- 762
 Qy 616 NSRDDGIFL-----LMPK-----HSANLW-----TTYQ-----VTPELTIG 646
 Db 763 DIRADRTFVTSYLFDAVQPSRYVLGLGYDHPDGIWGTINTMFTYSKAKSVDELGSQALLN 822
 Qy 647 GGVNAMSGITSSAGMHAGGYATFDAMAAYRFTPKLKLQINADNIFNRHY----- 695
 Db 823 GNANA-----KKAASRRTRPWYTVDSGVYINIKKGLTLRAGVYNLNLRYVTWENVROTAG 878
 Qy 696 -----YARVGGANTFNIPGSERTWTANLRY 720
 Db 879 GAVNQHKNVGVYNYAAPGRNYTFSELMKF 908

RESULT 42

US-08-337-483-94
 ; Sequence 94, Application US/08337483
 ; Patent No. 5923562
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney

```

STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-410 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-337-483-94

```

```

Query Match 4.4%; Score 167.5; DB 2; Length 908;
Best Local Similarity 18.2%; Pred. No. 1.8e-06;
Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

```

```

QY 3 QEFWSVPRIMTAATVLAALSSVFA-----AQTADLETVHLKGGORSYAIVTEKXGDY 55
DB 2 QOQHLEFLMLICSLMTAL--PYAENVQAEJOEKOLDITQYKAKOKTRRNEVTG-- 57
QY 56 SFEAVTVGTRIPASLREIPQSVSIITNOQYKRNVDFTDQARKEPGLRYLSDNDGSSV 115
DB 58 -----LGLTVKSSDTLSKEQVLN-----IRDLTYVDGIAVAVEGGRGASSG 98
QY 116 YA-RGYEYSHYN--IDGLPAQWOS-----INGTLP-----NLPAFDR--VEVWRCP 156
DB 99 YSIRGMDKRRVSLTVDGV--SQIQSYTAQALGGTRTAGSGAINEIEYENVKAVEISKGS 157
QY 157 SGLFSSGEMGGLVN-----171
DB 158 NSEBYNGALAGVAFOPTTAADIIIEGKQWGIQSKTAYSGKDHALTOSIALAGSGGAE 217
QY 172 ---LVRRPTKAFQGHAAAGFTH-----KQYK-----AEADVSG 203
DB 218 ALLIYTKRRGREIHAHKDNGKGVQSFNRLVLDBKKEGSGQYRFVIEECHNGYAAKN 277
QY 204 SLNSDGSV---RRVMAQTVGASPR---PAE-----KNRHH-----233
DB 278 KLEDEASVDERKTVTQDYTGSRNLANPLEYSGOSWLFRRPGMHLDNRHYGVALERTQ 337
QY 234 ETF-----YAADWDINP-----DTVIGAGY-----LY 256
DB 338 QTFDTDMTVPAVFTSB-DVYPSLKLGLKGYSGDNKAERLFVQEGSGTLOGIGTGCVFY 396
QY 257 QORHLAPYNGLPADANNKLPSSLPOHVFVGADWNKFKKNSHSHVADLKHVFGNGGYGVGM 316
DB 397 DEHRTKRRYGV-----EYVYHND-----KDTWAD-----YARLSYDRGGI 432
QY 317 RYSDRDADSNYAAGSKLGMKTPAGRGPCNTADDKACAVGLGEIHKOKALAPASYSRPF 376
DB 433 DLDNRLQOHTCHSDGSKNCRPDGNKPYSFYKSDRMT-----YESSRNLFOAVFKAP 485
QY 377 RLGNLANEFYIGADYRFRSTNEGRTLLYAFGGLALNEFRSIPQVLLINAR---KGV 432
DB 486 DTAKIRHNLSINIGYDRFKS--QLSHSDYLLQ-----NAVOAYDLITPKKPPPPNGS 535

```

```

QY 433 RGVSHTV-----TENIDFCITGKSTFHPADGLSLIGG-----RLGHYKIESGEG 479
DB 536 KDNPFKRVISIGKTTVNTSPICRFGNNITYDCTPRNIGNGNGYAAVDPNVRLGRW-ADVAG 594
QY 480 -----TLHKASKTKRFTG-----YAGAV-----YDLNDNNSLYL---SLSLQTYTPQT 518
DB 595 IRDYRSTHSEDSKSVSTGHRNLSMNAGVVLKPFYMMDLTYASSTGFRLLPSFAEMGMFA 654
QY 519 NLDADEKLKPRQNGQFEVG--YKG-----SYMDDRINAEVSF-YRMKDKN-----AA 563
DB 655 GSKLTKDLKPKERSFRRREGIVFKDGFNLKLEASYPFNNAARDLJAPGEYRTNGQTSASG 714
QY 564 APLNPNKKTRVAA---LGRVMEGEVTEISGAVTPKW-----QIHAGYSYHSQIKTAS 615
DB 715 DPGYRNAQNAIAGIINILCKIDMHGV-----MGCLPDGAYSTLAIYRIKVKDA- 762
QY 616 NRDDGIFL-----LMRK-----HSANLW-----TTYQ-----VTPELTIG 646
DB 763 DLRADRTFTSYLPDAVQSRVYLGIDYHDPDGIWINTMFTYSKAKSVDELIGSQALLN 822
QY 647 GGVNAMSGITSSAGMHAGVATFDMAAAYRFTPKLKLQINADNIFNRHY-----AA 695
DB 823 GANA-----KKAASRRTRPWYVTVDSGYINIKKHLTLRAGVYNLNLNRYVTWENRQTAG 878
QY 696 ----YARVGANFENIPGSERTWTANLRY 720
DB 879 GAVNQHKNVGVNRYAPAPGRNYTFSLKMF 908

```

```

RESULT 43
US-08-478-373-94
Sequence 94, Application US/08478373
Patent No. 5922841

```

```

GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Mardin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,373
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973

```

REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-478-373-94

Query Match 4.4%; Score 167.5; DB 2; Length 908;
Best Local Similarity 18.2%; Pred. No. 1.8e-06;
Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

Qy 3 QPMSVFRINMTAATVLAALSSVFA-----AQTADLETVHIKQORSYNAIVTEKNGDY 55
Db 2 QOHLFRNLICLSLMTAL--PYIAENVQAEQKQDITQVAKKQKTRDRDNEVTG-- 57
Qy 56 SSFAVTGTGKIPASLREIPQSVSIITNOQVKDRNVDTFDQLARKTPGLRVLSNDDGRSSV 115
Db 58 -----LGLKSSDSTLSKEQVLN-----IRDLTRYDPGIAVVEQGRGASSG 98
Qy 116 YA-RGYEYSEYN--IDGLPAQWQS-----INGTLP-----NLPAFDR-----VEVMRGP 156
Db 99 YSIRGMDKNRVSLTVDGV--SQIQSVTAQAALGGTTRTAGSSGAINEIEYENKVAIVEISKGS 157
Qy 157 SGLFDSSGSGMGGIN-----SGLFDSSGSGMGGIN-----SGLFDSSGSGMGGIN 171
Db 158 NSSEYVNGALAGSVAFTQTTADITGEGKQWGIQSKTAYSGKDHALTOSLALAGSSGAE 217
Qy 172 ---LVKRPTKAFQGHAAAGFGTH-----KQYK-----ABEADYSG 203
Db 218 ALLIYTKRGREIHAHKDAGKGVQSNRLVDBDKKGGSQRYFIVEECHNGYAAACKN 277
Qy 204 SLNSGQSV---RGRVMAQTVGASPR---PAB-----KNNRH----- 233
Db 278 KLKEDASVKDRKTVSTQDYTGNSRLANPLEYSGQSWLFRPGWHLDRHRYVGAVLERTQ 337
Qy 234 ETF-----YAADWDINP-----DTVLGAGY---LY 256
Db 338 QTFDTRDMTVPAYFTSE--DYVPGSLKGLKYGSDNKAERLFVQGGSTLQIGYGTGVFY 396
Qy 257 QORHLAPYNGLPADANNKLPSPQHFVFGADWNKPKMNSHDFADLKHYFGNGGKRYGM 316
Db 397 DERHTKNRYGV-----EYVYHNAD-----KDTWAD-----YARLSYDRQGI 432
Qy 317 RYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 376
Db 433 DLDRNLQOQTHCHSDGSDKNCRDPGNKPKSYFYKSDRMI-----YEESRNLPAVFKKAF 485
Qy 377 RLGNATANEPVIGADYNRPRSTNEQGRRTLYARGGLALNEFRSIPQVDLIANAR----KGV 432
Db 486 DPAKTRHNLINSLNGYDRFKS--QLSHSDYLLQ-----NAVQAYDLITPKPPFPNGS 535
Qy 433 RGYSHTV-----TENLDEFGYKSTFFHPADGLSLIGG-----RLGHYKIESGEG 479
Db 536 KNPYRVISGKTTVNTSPICREGNNTYDCTPRNIGNGGYAAVQDNVRLGRW-ADVAG 594
Qy 480 -----KTLHKASKTKPTG-----YAGAV-----YDLNDNNSLYL---SLSQLYTPTQ 518
Db 595 IRYDYRSTHSEKSVSTGTHRNLNAGVVLKPFPTWMDLTYRASTGFRPLPSFAEMYGWRA 654
Qy 519 NLDAADGKLLKPGQNOFVG--YKG-----SYMDRLNARVSP-YRMKDKN-----AA 563
Db 655 GESLTKLDLKPESKFNREAGIVFKGDFGNLEASYFNNAIRDLIAFGYETRTQNGQTSAG 714
Qy 564 AFLPNNNKTRVAA---LGRKRVMEGVETEISGAVTPKW-----QIHAGVSYLHLSQIKTAS 615
Db 715 DPGYRNAQNAIAGINILGKIDWHGV-----WGGLPDGLYSTLAYNRKIVKDA- 762
Qy 616 NSRDDGIFL-----LMPK-----HSANLW-----TTYQ-----VTTELITG 646

Db 763 DIRADRTFVTSYLFDAVQPSRYVLGLGDYHDHDTGWTMTFTYSKAKSVDELGSQALLN 822
Qy 647 GGVNAMSGITSSAGMHAGGYATFDAMAAYRFTPKLKLQINADNIFNRHY-----695
Db 823 GNANA-----KKAASRRTRPMVYTVDSGVYINIKKHLTLRAGVYNLLNYRYVTWENVROTAG 878
Qy 696 -----YARVGANTFNIPGSERTWTANLRY 720
Db 879 GAVNQHKNVGVYNYRYAAGPRNYTFSELMKF 908

RESULT 44
US-08-474-671-94
Sequence 94, Application US/08474671
Patent No. 6008326
GENERAL INFORMATION:
APPLICANT: Loesmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,671
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-474-671-94

Query Match 4.4%; Score 167.5; DB 3; Length 908;
Best Local Similarity 18.2%; Pred. No. 1.8e-06;
Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

Qy 3 QPMSVFRINMTAATVLAALSSVFA-----AQTADLETVHIKQORSYNAIVTEKNGDY 55

Search completed: December 18, 2002, 06:47:00
Job time : 28.862 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:40:34 ; Search time 27.0814 Seconds
(without alignments)
2562.980 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776

Sequence: 1 MGQFMSVFRINMTATVLA.....NTFNIGSERTWTNLRYSF 722

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3244	85.9	708	2 B81038	TonB-dependent rec
2	832	22.0	802	2 A83125	probable TonB-depe
3	828.5	21.9	826	2 S41569	ferric pseudobacti
4	822.5	21.8	813	2 A40601	ferripyoverdine re
5	821.5	21.8	815	2 H83345	ferripyoverdine re
6	788	20.9	725	2 D81976	probable ferric si
7	784	20.8	725	2 H81030	TonB-dependent rec
8	767.5	20.3	819	2 S15169	ferric-pseudobacti
9	755	20.0	753	2 A80187	probable iron-side
10	737.5	19.5	729	2 C64854	ferric-coprogen re
11	736.5	19.5	729	2 H90813	outer membrane rec
12	734.5	19.5	729	2 D85673	outer membrane rec
13	715.5	18.9	720	2 A36942	Fe(III)-pyochelin
14	708.5	18.8	809	2 S32899	ferric-pseudobacti
15	512	13.6	863	2 AF2074	ferrichrome-iron r
16	486.5	12.9	820	2 AE2130	ferrichrome-iron r
17	486.5	12.9	885	2 AB1944	ferrichrome-iron r
18	483.5	12.8	854	2 AE2082	ferrichrome-iron r
19	481	12.7	858	2 AD2421	ferrichrome-iron r
20	479	12.7	851	2 D87252	TonB-dependent rec
21	475	12.6	863	2 A12077	ferrichrome-iron r
22	470	12.4	760	2 E64817	probable membrane
23	467.5	12.4	867	2 AC2140	ferrichrome-iron r
24	465	12.3	828	2 S74450	ferrichrome-iron r
25	460.5	12.2	851	2 AE2129	ferrichrome-iron r
26	460	12.2	760	2 C90739	hypothetical prote
27	460	12.2	760	2 E85589	hypothetical prote
28	459	12.2	708	2 A83043	probable outer mem
29	450.5	11.9	747	2 G97408	fegA protein U6140

30	450	11.9	819	2 A13197	TonB-dependent rec
31	449.5	11.9	872	2 AC2134	ferrichrome-iron r
32	441.5	11.7	708	2 AH2626	ferrichrome-iron r
33	441.5	11.7	753	2 D83081	probable outer mem
34	433.5	11.5	732	2 A83481	probable TonB-depe
35	424	11.2	858	2 AE2085	ferrichrome-iron r
36	421.5	11.2	696	2 AC0547	ferrichrome-B rec
37	420.5	11.1	853	2 AC2079	ferrichrome-iron r
38	416.5	11.0	853	2 S74457	ferrichrome-iron r
39	415.5	11.0	857	2 AC2132	ferrichrome-iron r
40	411	10.9	714	2 C95382	probable ferrichro
41	407.5	10.8	635	2 C81861	hypothetical prote
42	402	10.6	863	2 S74447	ferrichrome-iron r
43	396.5	10.5	710	2 S22673	ferrichrome recep
44	396.5	10.5	802	2 C83588	probable hydroxama
45	391	10.4	713	2 E91118	probable ferrichro

ALIGNMENTS

RESULT 1

TonB-dependent receptor NMB1829 [imported] - Neisseria meningitidis (strain MC58 serogro
C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C/Accession: B81038
R/Tetrelm, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Zi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A/Reference number: AB1000; MUID:20175755; PMID:10710307
A/Accession: B81038
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-708 <TEXT>
A/Cross-references: GB:AE002532; GB:AE002098; NID:G7227078; PIDN:AAF42164.1; PID:G722708
A/Experimental source: serogroup B, strain MC58
A/Genetics:
A/Gene: NMB1829

Query Match	Score 3244; DB 2; Length 708;
Best Local Similarity 86.8%; Pred. No. 6.1e-219;	
Matches 627; Conservative 33; Mismatches 48; Indels 14; Gaps 3;	
QY 1 MGQFMSVFRINMTATVLAALSSVFAAQTADLETVHIGQSYNAIVTEKNGDYSSFAV 60	
DB 1 MGQFMSVFRINMTATVLAALSSVFAAQTADLETVHIGQSYNAIVTEKNGDYSSFAV 60	
QY 61 TYGTRKPAALREIPQSVSIITNOQVDRVDFPDQARPTGLRVLSNDGSSVARYQ 120	
DB 61 TYGTRKPAALREIPQSVSIITNOQVDRVDFPDQARPTGLRVLSNDGSSVARYQ 120	
QY 121 EYSEYVNDLPAQWOSINGTLPNLFAFDRVEYMRGSGHFDSSGEGGIVNVRKPTPA 180	
DB 121 EYSEYVNDLPAQWOSINGTLPNLFAFDRVEYMRGSGHFDSSGEGGIVNVRKPTPA 180	
QY 181 FQHAAGAAGFTGHQYKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEKNNRRETFYAAA 240	
DB 181 FQHAAGAAGFTGHQYKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEKNNRRETFYAAA 240	
QY 241 DMDINDPTVLGAGYLYQOCHLAPYNGLPADANNKLPSPLOHVFVGADWKKFKFNHSDVFA 300	
DB 241 DMDINDPTVLGAGYLYQOCHLAPYNGLPADANNKLPSPLOHVFVGADWKKFKFNHSDVFA 300	
QY 301 DLKHYGNGGYGVGRVYRDRDADSNVAFAGSKLGWKTAPGPGCNTADDKACAVGLGTE 360	
DB 301 DLKHYGNGGYGVGRVYRDRDADSNVAFAGSKLGWKTAPGPGCNTADDKACAVGLGTE 360	
QY 361 IYKAKALFPAASYSRPRLCNTANEFYTGADYVFRSTNEQSTTLVYARGGLALNERSLP 420	

Db 349 IKQKAFVADASYRPFALGNTANEFVIGADYNLRSTNEQGRSTL--SKSVALDGFRLP 406
Qy 421 QVDLIANARKGVGRVSHVTATENLDEFGYIGKSTHPADGLSLGGRLGHYKIESGDK 480
Db 407 YNGILQARAGNAGNFHNSVTENLDETLGYAKTFVFRPLEGLSLIAGGRVGHKIESGDK 466
Qy 481 TLHKASKTKFTGYAGAVYDLNNSLYLSLSOLYTPQTNLDADGKLLKPRQGNQFEVGYK 540
Db 467 TLHKASKTKFTGYAGAVYDIDGNSLYASASQLYTPQTSIGTDGKLLKPREGNQFEIGYK 526
Qy 541 GSYMDRLNARYSFYRMKDKNAAAPLNPNKKTRVAALGKRVMEGVETEISGAVTPKQOI 600
Db 527 GSYMDRLNTRVSFYRMKDKNAAAPLDSNNKTRVAALGKRVMEGVETEISGAVTPKQOI 586
Qy 601 HAGSYLSHQIKTASNRDGDIFLLMPKHSANLWTTQVTPELTIGGGVNAMSGITSSAG 660
Db 587 HAGSYLSHQIKTASNRDEGIFLLMPKHSANLWTTQVTPSLTIGGGVNAMSGITSSAG 646
Qy 661 MHAGGYATFDMAAAYRFTPKLQINADNIFNRHHYARVGGANTNIPGSERTWTANLRY 720
Db 647 IHAGGYATFDMAAAYRFTPKLQINADNIFNRHHYARVGSSESTNIPGSERTWTANLRY 706
Qy 721 SF 722
Db 707 SF 708
RESULT 2
A:Title: TonB-dependent receptor PA4168 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83125
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; UID:20437337; PMID:10984043
C:Accession: A83125
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <STO>
A:Cross-references: GB:AE004833; GB:AE004091; NID:g9950370; PIDN:AAG07555.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4168
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homo
Query Match 22.0%; Score 832; DB 2; Length 802;
Best Local Similarity 30.0%; Pred. No. 4.5e-50;
Matches 222; Conservative 137; Mismatches 309; Indels 72; Gaps 21;
Qy 19 AALSSSVFAAQADLETVHIKG-QRSYNAIVTEKNGDYSSFAVTVGTIKIPASLRIPQSV 77
Db 97 AATGAGIAPGSDVLDGALSVATQDPATAITEDSGSYTTRAMRTSTKLAMSIRETPQSV 156
Qy 78 SIITNQVKDRNVDFDQIARLTPGLRVLNSDDGRSSVYARGYEYSEYNIDGLPAQMOSI 137
Db 157 SVVTRQRMDDQGRDLNDAVKGVGTGLTVQYGPARKVYAARGFDVDNIMYDGLPISISTY 216
Qy 138 NG---TLNPLFAFDRVEMVRGSGFLFDSGSGGGIUNLVKRPYKAFQGHAAAGGTHKQ 194
Db 217 TDQVISAADLAMDREVVVRGATGLMQAGNPAIAINMVRKRPQTEFRASLQGSVGTWDR 276
Qy 195 YKAEADVSGSLNSDGSVGR-VMAQTGASPRPAEKNRHETFYAADWDINPDITVLGAG 253
Db 277 YKAEADVSGSLNSDGLRGAIVAQKGNPSQDHVSSER-GVFGIARADUSPDPTFAIG 335
Qy 254 YLYQORHLAP-YNGLPADANNKLPSPQHFVFGADWNKFKMNSHDVADFADLKHFGNGGYG 312
Db 336 ASNQNDNRNDNVWGLPSPLGGRHLDLKRSSYYGADSWYDWDITTHLFSDLTHRFANGWQM 395

Qy 313 KVMRYSDRDADSNYAFAG-----SKLGMKTPAGRPG-CNTADDKACAVGLGTEI 361
Db 396 KLA-----ADKLWAINMLGLYNDVCYISTGCSMTQNPGDYSYTDHHD----- 439
Qy 362 KQKALAFDASYSRFRICNTANEFVIGADYNFRSTNEQGRRTLLYARGGL---ALNEFRS 418
Db 440 -----SDVAYANGFPQLLGRHEHLVVGASVROERFDGHHGWSLFNKDGTPTGMDPTQM 494
Qy 419 IPQVDLIANARKGVGRVSHVTATENLDEFGYIGKSTHPADGLSLGGRLGHYKIESGE 478
Db 495 DPSSTLKPRLNTSLWGMK-----LDQEQKGYALTTRNLADPLKVLGGRLDWMYKADAT 549
Qy 479 KTLHKASKTKFTGYAGAVYDLNNSLYLSLSOLYTPQTNLDADGKLLKPRQGNQFEV 538
Db 550 DS--YKVTN-NTVTRVAGVIYDLNQTSYVASYTDIFKPSNFDAGGGLDITKQNYEIG 606
Qy 539 YKGYMDRLNARYSFYRMKDKNAA-----APLNPNKKTRYA--ALGKRVMEGVE 587
Db 607 LKGFHFGALNSQIALFOIDENRATEDVGGPSPCPSPPT---SRYCSRAGSKVRSQVD 663
Qy 588 TEISGAVTPKQIHAQSYLSHSLQIKTASNRDGD--IFLLMPKHSANLWTTQVTPEL-- 643
Db 664 LELSGALSDDWQWAGVTVYDAKYKHSNKAKEGKPPDAAKPRHLFKLATSYTLPGLHLK 723
Qy 644 -TIGGGVNAMSGI-TSSAGMHAGGYATFDMAAAYRFTPKLQINADNIFNRHHYARVGG 701
Db 724 WRVGGDLATOSKTESDSTGFGQGGYTVVNAMLYKVNERIDTRLNPNLFDKYYYSIDF 783
Qy 702 ANTENIPGSERTWTANLRY 721
Db 784 GN-LNY-GEPRNLMTVKYS 801
RESULT 3
S41569
ferric pseudobactin M114 receptor protein - Pseudomonas sp.
C:Species: Pseudomonas sp.
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S41569; S34357
R:Morris, J.; Donnelly, D.F.; O'Neill, E.; McConnell, F.; O'Garra, F.
Mol. Gen. Genet. 242, 9-16, 1994
A:Title: Nucleotide sequence analysis and potential environmental distribution of a ferr
A:Reference number: S41569; UID:94104606; PMID:8277948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-826 <MOR>
A:Cross-references: EMBL:X73412; NID:g313725; PIDN:CAA51812.1; PID:g313726
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homo
F:183-317/Domain: tonB-dependent receptor carboxyl-terminal homology <TNN>
F:540-826/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
Query Match 21.9%; Score 828.5; DB 2; Length 826;
Best Local Similarity 29.9%; Pred. No. 8.4e-50;
Matches 227; Conservative 129; Mismatches 303; Indels 101; Gaps 26;
Qy 16 TVLAALSSSVFAAQADLETVHIKGQRSYNAIVTEKNGDYSSFAVTVGTIKIPASLRIPQ 75
Db 115 TVLGPATGSAM-----ELAPTNVNASRL--GATTEGNSYTTGGVTIGKV-HSLKETPQ 166
Qy 76 SVSIITNQVKDRNVDFDQIARLTPGLRVLNSDDGRSSVYARGYEYS-BYNIDGLP--- 131
Db 167 SVTVTRKMLDDQNLNLTIEQVMEKTPGITVYDSPMGKGYFSRGRMSGGQYQYDGPFLDI 226
Qy 132 ----AQMOSINGTLNPLFAFDRVEMVRGSGFLFDSGSGGGIUNLVKRPYKAFQGHAAA 187
Db 227 GSSVYQADSPNS---DMAIYDRVEVLRAAGMMKAGAGTAGGVNFRKRGQDTHAHTLSL 283
Qy 188 GFGTHKQYKAEADVSGSLNSDGSVGR--VMAQT-----VGASPRPAEKNRHETFYAA 239
Db 284 SAGTWDNRYGQVDTGGFLNSDGTIRGAVTEQTRQYFYDVG-----SRKQDIYYGA 335

A:Residues: 1-815 <STO>
A:Cross-references: GB:AE004666; GB:AE004091; NID:g9948438; PIDN:AA05786.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:

A:Gene: fpvA; PA2398
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom

Query Match 21.8%; Score 821.5; DB 2; Length 815;
Best Local Similarity 29.0%; Pred. No. 2.5e-49;
Matches 217; Conservative 145; Mismatches 312; Indels 75; Gaps 24;

```
QY 8 PRINNTAATVLAALSSVFAAQTADLETVHIKQORSYNAIVTEKNDYSSFAVTVGTKIP 67
DB 108 FCGNAITTSVAEADSSV-----DLGATMITSNQL--GTITEDSGSYTPGTIATRLV 159
QY 68 ASLRIPQSVSIITNQVKDRNVDPDLARKTPGLRVLSNDDGRSSVYARVEYSEYNI 127
DB 160 LTPRETQSITVTTRQNMDDFGLNIDDMVHTPGITVSAYTDRNNYARFSSINNFOY 219
QY 128 DGLPAQMSI-----NGTLNLPFAFVRVEMRPGSLFDSSGEMGIVNLVRKPTKAFQG 183
DB 220 DGIPSTARNVGVSGATLSMAIYRVEVLKATGATGLLTGAGSLGATINLRKKPTHEFKG 279
QY 184 HAAAGFGTHKQYKAEADVSGSLNSDGSVRGVMAQTVGASPRPAEKNNRHETFYAADWD 243
DB 280 HVELGAGSDNYRSELDVSGPTESGNVRGRAVAAAYQDKHSFMDHYERTSVYYGILEFD 339
QY 244 INPDTVLGAGLYQQRHL--APYNG-LPA-DANNKLPLSPQHFVFGADWNKFKMNSHDVF 299
DB 340 LNPDTMLTVGADYQNDPKSGWSGSPFLFDQGNRNDVRSFNNGAKWSSHEQYTRTVF 399
QY 300 ADLKHFGNGGYGKVMRYSDRDAVSNYAFAGSKLGMKTPAGRPCNCTADDKACAVGLGT 359
DB 400 ANLEHNFGWGVKQVL--DHKINGYHAPLGAIMG-DWPA-----PDNSAKIVAAQKY 448
QY 360 EIKQALAFDASYSRPRFLGNTANEFVG--ADYNRFRSTNEQGRITLYARG-GLALNEF 416
DB 449 TGGETKSNLSLDIYLPQFQFLGREHELWGTSAFSHW-----EGKSYWNLRYNDNTDDF 503
QY 417 RSIPQVDLIANARKGVRGVSHTVATENDEFGIYKGSTPHPADGLSLGGGRGLGHYKTES 476
DB 504 -----INWDGDIKDPDWGTPSQYIDKTRQLGSYMTARENVTDNLNLFGRVVDYRV-T 557
QY 477 GEGKTLHKASKTKFTGYAGAVYDLNNSLYLSLSOLYTPQTN--LDADGKLLKPRQNG 534
DB 558 GLNPTIRESG--RFIPYVGAVIDLNDTSVYASYTDIFMPQDSWYRDSNKLLEPDEGON 615
QY 535 FEVGYKGSYMDRLNARVSFYRMKQNAAP-----LNPNNKKTYYAALG-KRVMEGVET 588
DB 616 YEIGIKGEYLDGRLNTSLAYFEIHEENRABEDALYNSKPTNPAITYAYKIKAKTKGYEA 675
QY 589 EISGAVTPKQWIHAGSYLHLSQIKTASNRDDGIFLLMPKHSANLWTVYQ---VTPELTI 645
DB 676 EISGELAPGWQVAGYT--HKIIRDDSGKK---VSTWEPQDQLSLYTSYKFKGALDKLTV 730
QY 646 GGGV-----NAMSGITSSAGMHAGYATFDMAAAYRFTPKLQLOINADNINFR 693
DB 731 GGGARWQGSQWQVNNPRSRWEK---FSQEDYVLVDLMARYQITDKLSASVNVNVFDK 787
QY 694 HYARVGANTENIPGSERTWTANLYSF 722
DB 788 TTYTNIIGFYSASY-GDPRNLMFSTRWDF 815
```

RESULT 6
DB1976
probable ferric siderophore receptor protein NMA0575 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: DB1976
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: DB1976
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83866.1; PID:g737931
A:Experimental source: serogroup A, strain Z2491
C:Genetics:

A:Gene: NMA0575
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homo

Query Match 20.9%; Score 788; DB 2; Length 725;
Best Local Similarity 29.9%; Pred. No. 4.7e-47;
Matches 232; Conservative 117; Mismatches 316; Indels 112; Gaps 23;

```
QY 5 MSVFRINWTAATVLI-----AALSSSVFAAQTADLETVHIKQORSYNAIVTEKNGDYSS 57
DB 1 MTRFKYLSLFAALLPVYQAQDVSDDPKQESTELPTITVTADRT-----ASSNDGTVV 55
QY 58 FAVTVGTKIPASLREIPQSVSIITNQVKDRNVDPDQ-LARKTPGLRVLSNDD--GRSS 114
DB 56 SGHTPLGLPMTLREIPQSVSVIITSQQMRDQNIKALDRALLQAATGTSRQIYGSBRAGVY 115
QY 115 VYARGYEYSEYNIQGLPAQMSQINGTLPNLPFDRVEMRPGSLFDSSGEMGIVNLVR 174
DB 116 LFARGSRANYQINGIPVADALADTGNANTAAAYERVEVRGVAGLLDGTGEPSPATNLVR 175
QY 175 KRPTKAFQGHAAAGFGTHKQYKAEADVSGSLNSDGSVRGVMAQ--TVGASPRPAEKNNRH 233
DB 176 KRPTKPLFEVRAAGNRKHPGLGADVSGSLNAEGTLRGRUVSTFGRGDSWRQER--SRD 234
QY 234 ETFYAAADMDINPDTVLGAGLYQQRHL---APYNGLPADANNKLPLSPQHFVFGADWNK 290
DB 235 AELYGILEYDIAPQTRVHAGWDYQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWAN 294
QY 291 FKMNSHDVADLKHYPNGGKGVGMRYSDRDAVSNYAFAGSKLGMKTPAGRPCNCTADD 350
DB 295 SRHALNLFAGIEHRFN-----QDWKLKAEYDYTRSR--FRQPYGAVGVLSDIH 341
QY 351 KACAVGL--GTEIKQKALAFDASYSRPRFLGNTANEFVGADYNRFRSTNEQGRITLYA 407
DB 342 NTAATDLIPGVWHADPRTHSASVSLIKYRLFGRHDLIAGINKYKAS-NKYGE----- 395
QY 408 RGLALNEFRSIPQVDLIANARKGVRGVSHTV-----TENLDEFGI-----YKST 454
DB 396 -----RSI-----IPNAIPNAYEFSTRGAYPQPASPAQTIPOYGTRRQIGGYLATR 441
QY 455 PHPADGLSLICGGRGLGHYKTESGEGKT--LHKASKTKFTGYAGAVYDLNNSLYLSLQ 512
DB 442 FRAADNLSLILGGYSRYRTGSDYSTRGQMTYVSANRFTPTGTGVFDLTGNLSLYGSYS 501
QY 513 LYTPTQTNLDADGKLLKPRQNGQFVGKSGSYMDRLNARVSFYRMKDKN--AAAPLNPN 570
DB 502 LFVQSQKDEHSGVLYKPVGTGNNEAGIKGEWLEGRNLASAAYVARKNNLATAAGRDPSG 561
QY 571 KKTRYAALGRKVMGEVETESGAVTPKQWIHAGSYLHLSQIKTASNRDDGIFLLMPKHS 630
DB 562 -NTYRAANQAKTHWEIEVGGRIITPEWQIAGYSQSKTRDQDGSRLNPDSV----PERS 616
QY 631 ANLWTTTQVTPEL-----TIGGV-----NAMSGITSSAGMHAGYATFD 670
DB 617 FKLFTAYHFAEPAPSGWTIGAGVRWQSETHDTPATLRPNPAAKARAADNSRQKAYAVAD 676
QY 671 AMAAYRFTPKLQLOINADNINRHHYARVGGANTFNIP-----GSERTWTANLYSF 722
DB 677 IMARYRNPRAELSLNVDNLFNKHRYTQ-----PDRHSYGALRTVNAAFYRFP 724
```

RESULT 7
H81030
TonB-dependent receptor NMB1882 [imported] - Neisseria meningitidis (strain MC58 serogro
C:Species: Neisseria meningitidis

C|Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C|Accession: H81030
R|Etcellin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.P., Hickey, E.K., Hafe, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., H., Qin, H., Vamathavan, J., Gill, J., Scarlato, V., Massignani, V., Piazza, M. Science 287, 1809-1815, 2000
A|Authors: Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, B.R., Rappuoli, R., et al. |Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. |Reference number: A81000; MUID:20175755; PMID:10710307
A|Accession: H81030
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-725 <Tern>
A|Cross-references: GB:A0002538, GB:AE002098, NID:g7227136, PIDN:AAF42216.1; PID:g7227136
A|Experimental source: serogroup B, strain MC58
A|Genetics:
A|Gene: NMB882
A|Superfamily: ferric-pseudobactin receptor; tonb-dependent receptor amino-terminal homo

```

Query Match          20.8%; Score 784; DB 2; Length 725;
Best Local Similarity 29.4%; Pred. No. 8.9e-47;
Matches 228; Conservative 118; Mismatches 320; Indels 110; Gaps 21;

OY MSVFINTMTAAVL-----AALSSVPAQTALFVTHIGORSNAVALTEKKGDYSS 57
Db 1 MTRFKYSLFPAALLVYVQADVSDDPKPEQESTELPITTVADRT-----ASSNDGYTV 55
OY 58 FAVTVGTRKIPASLREIPQSVSIITNQOVKDRNVDFDQ-LAKTGTGLRYLSND--GRSS 114
Db 56 SGTHTPLGLPMTLRREIPQSVSVITSQMRDQNIKTLDRALLQATSTSHQIYSSDAQAGVNY 115
OY 115 VYARGIYESEINKIDGLPAMQOSINGTLPNLFADFVEVMKSGSLPDSGEMKGIVNIVR 174
Db 116 LFARCSRLANYQINGI PVADALADTGMNNTAAVEEVEVVRGVAGHLLDGTGSPSATVTVNR 175
OY 175 KRPTAFQGHAAAGGTGHQYKAEADVSGSLNSDQSVGRVVAQTVGASPRPAEKNNHE 234
Db 176 KRLTKKPLPEVRAEAGNRHGFGLDADVSGSLNTEGTLKRLVLTSTGRQDSWRERRSRDA 235
OY 235 TFYAADDNDINDPTVLGAGYLYLQQRHL---APYNGLPADANNKLPSLPQHVEVAGDMKPF 291
Db 236 ELYGLEDIDIAQOTRVHAGMDYQQAKEPTADAPLSAAYVDSQGYATAFPGKPNPATNMANS 295
OY 292 KMNSHDVFADLKHVFGNGGYGVKMKRYSDDRADSNYAFAGSKLGMKTPAGRGCTADDK 351
Db 296 RHRALNLFAGIEHRFN-----QDWKKAEDYDTRSR--PRQPGVAGVLSIDHN 342
OY 352 ACAVGL---GTIEKQKALAFDASYSRPFPLGNTANEFYVIGADYNFRSTNEGGRTTLVAR 408
Db 343 TAAATLIDPYWHADPRTSHASVSLIGKYRLPGRHDLIAGINGVYVAS-NKYGE----- 395
OY 409 GGLALNEFSSIPQVOLDINARKGVGRYSHTVA-----TENDLDFRGI-----YKSNPF 455
Db 396 -----RSL-----IPNALPNAVEFSEFTGAYPOPASPAQITIPQGTIRROIGGYLATRF 442
OY 456 HPADGLSLIGGRLGHYKYLESGEGKT--LHKASKTKFTGYAGAAVVDLNDNNSLYLSLSOL 513
Db 443 RAADNLSTLIGGRYTRRGRYSYDSRTQGTYYSANRFPYTGIVVDLIGNLSLYSSYSYL 502
OY 514 YTPQTNLADGKLLKPRQGNQPEVVGKGSYMDRLNARVSFTRMKDKN--AAAPLNPNK 571
Db 503 FVPOQKQEHGSGYLLPVPVGNNLLEAGIKGEMLEGRINLASAAVYRARKNMILATAGDRDPG- 561
OY 572 KRYVALGRVWEGVETESGAVTEPKMOIHAGYVLIHQIKTANSRDGIGLILMPKISA 631
Db 562 NTYYRAANDAKTHGWEIEVGRIETPMQIQAGYSQSKTDQCGSLRNLPVSV---PERSEF 617
OY 632 NLMTTYQVTVPEL-----TIGGGV-----NAMSGITSSAGMHAGGATFPA 671
Db 618 KLFTHVHFAPEARPSGWTIGAGVROMSETHTDPATILIRIPRAKAAARADSRQKATAVADI 677
OY 672 MAAYRTPKLLQIANDNIENRHYARVGAGANTENIP-----GSERTWTANLRYSGF 722

```

RESULT 8

S15169
ferric-pseudobactin receptor precursor - Pseudomonas putida
C|Species: Pseudomonas putida
C|Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 29-Sep-1999
C|Accession: S15169
R|Balter, W.; Marangy, J.D.; de Weger, L.A.; Tommassen, J.; Weisbeek, P.J.
Mol. Microbiol. 5, 647-655, 1991
A|Title: The ferric-pseudobactin receptor Pupa of Pseudomonas putida WCS358: homology to
A|Reference number: S15169; MUID:91260449; PMID:1646376
A|Accession: S15169
A|Molecule type: DNA
A|Residues: 1-819 <Bit>
A|Cross-references: EMBL:X56605; NID:g45722; PIDN:CMA9942.1; PID:g45723
C|Genetic8:
A|Gene: pupa
C|Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homo
C|Keywords: membrane protein
F|1-47/Domin: signal sequence #stratus predicted <SIG>
F|48-819/Product: ferric-pseudobactin receptor #stratus predicted <MAT>
F|189-322/Domai: tonB-dependent receptor amino-terminal homology <TN>
F|542-819/Domai: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 20.3%; Score 767.5; DB 2; Length 819;
Best Local Similarity 27.8%; Pred. No. 1,56-45;
Matches 214; Conservative 134; Mismatches 306; Indels 117; Gaps 26;

D 8 PRINATATVLAALSSVEAAQTADLETTHIKGRSYNAIVTEKNDYSSFAVTGKIP 67
::: ||||| :::: |::| :|::| :|::|
D 110 YQIDENTVTVASA-----AAKDQIELSATNVNSAGIETGETGTSTATTSTATKM 164
::: ||||| :::: |::| :|::| :|::|
OY ASIREIPSVSILITNOVKDKNNVDTPQLAKRTGLRLSDNG-RSSVARGVEYEYN 126
68 ASIREIPSVSILITNOVKDKNNVDTPQLAKRTGLRLSDNG-RSSVARGVEYEYN 126
D 165 LSIRTPPTITVVTRKQRDDHDLGSMEVLTYTGFI-TMSODGEREFTYSRGSAINIYO 223
::: ||||| :::: |::| :|::| :|::|
OY 127 IDGLPA----OMOSINGTLPMIFAFDRVENVRGSGSLFDSSGEMGAIYNLRKSPTKAFQ 182
D 224 FDGVTTYYDNQNRNPSTLMVGAYDRLEIVRGATGLMTGAGDSAVVNVRKKPTEFEK 283
::: ||||| :::: |::| :|::| :|::|
OY 183 GHAAAFGTHKOYKAADVSGLSINDSGSVRGVAQTGASPRPAEKNNR-----H 233
::: ||||| :::: |::| :|::| :|::|
D 284 SHIQGVGSWPDYRAEADVSGPLTDGVRGRFFA-----AKDNHTFMMDYTQDR 334
::: ||||| :::: |::| :|::| :|::|
OY 234 ETFFAADMDINPDTVLGAGLYQQRIAPYNGLP-----DANKKLPSLPQHVVGDAMN 289
::: ||||| :::: |::| :|::| :|::|
D 335 DVLGYGVADVADVTIVARFGI--DRQTYKVNGAPGPPIITYNQOPTFSRSTSSDARM- 390
::: ||||| :::: |::| :|::| :|::|
OY 290 KFKNMSSHVPFADLKHYGF-----NGGYKGVGMRSDDADSNAPAFSGSKLGMTTPAGR 343
::: ||||| :::: |::| :|::| :|::|
D 391 ----GYDDYTTNTYTFLEEQALHDWOFKLAAMYMDVDRSSFSY-----YSTTNRS 439
::: ||||| :::: |::| :|::| :|::|
OY 344 GCNTADKACAVGLTEIKOKALAFDASYSRPFLIGNTANEFVIADYNERSTRNEOGRT 403
::: ||||| :::: |::| :|::| :|::|
D 440 YLELDGSTSEISAGIYT-AKHQHQGVDAITLQGFOLLGQTHELLIVGNVYLEXNHK----- 493
::: ||||| :::: |::| :|::| :|::|
OY 404 TLVYAGGLALNFESIPVDL-----IANARKG---VRGYSHTVATENLDEFGIYGS 453
::: ||||| :::: |::| :|::| :|::|
D 494 ---RGD-----SGRPVINIFYDMWDNQTPKPGDEIIPIGYNISNR--QSQGYFVAS 539
::: ||||| :::: |::| :|::| :|::|
OY 454 TPHPADGLSLIGGGGLGHYKIE-----SGEGKTLHKASKTKFYGUYGAVALDLNDNNSLY 507
::: ||||| :::: |::| :|::| :|::|
D 540 RNLTDJLHLILGARASNYRFDYALMRI GNBPARYKVERGVTPYAGIYIDLTNBSOV 599
::: ||||| :::: |::| :|::| :|::|
OY 508 LSIQLYTPQTNLDADGKLLKPRQNGPEVYGKSYMDRINATAVSFYRMKDKXAAAPLN 567
::: ||||| :::: |::| :|::| :|::|
D 600 ASYTDIFPKQNNDITGKPLDPVGVKNVELGMKEPFLEGRIANNMIALYMVRDLASTN 659
::: ||||| :::: |::| :|::| :|::|
OY 568 ---PNNKTRVLAALGKRVMEGEV-----EISGAVTPKMQIHAGSYLHSQIKTASNSRD 619
::: ||||| :::: |::| :|::| :|::|
D 660 EVWPVPSG-----GLASRYVDAETKGVDVLELGSVELPDMNVFTGYS--HTTEDADGKR- 712
::: ||||| :::: |::| :|::| :|::|

Best Local Similarity 27.3%; Pred. No. 1.6e-43;
Matches 211; Conservative 148; Mismatches 308; Indels 105; Gaps 26;

```

01 3 QEMSAFRLNMTAAVY-LAALSSVFAAQTADLETHIKGKRYNALIVRKNDYSSPAVT 61
02 11 QYALTTSLSLGLCTALLLPSSAFAAPATE-ETIVIBESAT---APPDENDYSVTSTS 66
03 62 VGTKIPASLREIPQSVSIITNQOVKDRNVDTFDOJARKTPGRLVLSNDGRSARVGE 121
04 67 AGTKMQMTQRDPQSVTVIVSQQRMEDQOLQTIQGEVMENTLIGISKQADSDRLALYISRGFQ 126
05 122 YSEYNIDGLPAQMOS--INGTLPNLFAFDREYVNRGPBGLEFDSSGEMGIVNLVRKSP 178
06 127 IDNVYWDGIPYTFEESRMNLGDALSMALFERVEYVRGATGLMTGTGNPSAALINVRKAT 186
07 179 -KAPGCHAAAGCTHKQYKAEADVSGSLNSDSVAGRVAAQVGA---SPRAEKONNHE 234
08 187 SREFPGDVSAEAGSNKKERYADLQSPLEDEKIKARI---VGGYQONDSWLDREYSEK 242
09 235 TFVAA-ADMDIIPDPVLGAGVLYOORHL--APYNGJPA-DANNKJPLPOHVFVADNNK 290
10 243 TTFSGIVADADGLDTLLSAGYEQKIDVNSPTWGLPFKMTDGSNSIDRAASTAPDWAY 302
11 291 FRKNSHDVPADIKHYFGNGGYGKVMGRYSDDRADSNT---AFAASKJGMKTP----- 339
12 303 NDKEINKVFMTLKQGFADTWQATLNAHSEVEFDSKMMYVDAYVANKADOMLVGPYSNYGP 362
13 340 ----AGRGCCNTADKACAVGLGTE-----IKQALAFDASYSRPF--LGNTINEF-- 385
14 363 GFDVYVGTTGMSNGKRRVADLDFADGSYELFGROHNLFFGGSYSKONNRYPFSSWNIIPD 422
15 386 VIGADYNEFRSTNEQGRTLLYARGGLALNEFRSIPQVOLIANAARKGVRSHTVAENID 445
16 423 EIGSPYN-FNG-----NFPQTD-----WSPOSIAQDDTT 450
17 446 EF-GIYGKSTFHPADGLSIIGGRLGHYKIESGEGKTLHKASKTYFTGYAGAVYDLNUN 504
18 451 HMKSLYAATRVTLADPLHLILGARATNMRVDT---LTVSMKNTHTTPAAGLVPIINDNW 506
19 505 SLVYLSQLYPTQNLMDADGKLLKPRQNGQFEVYKGSYMDRLNARVSFYMKRKNA-- 563
20 507 STVASYTSIFQONQRDSGKTLAPITGANNYELGAKSDMMSNRLTTTLTAIFRIEDONVAQ 566
21 564 ---ABLNENKKTTRVYALGKRVMEGVETEISGAVTPKQIHAGYS-YLHSQKLTASNSRD 619
22 567 STGTPIPGSNGEATAKADGVIVSKVEFELNAALITDNMQLTFTGATRYI-----AENDEG 620
23 620 DGIFFLMPKHSANLWTTYQ--VTPELLIGGVANMSGITSSA-----GMHAGYATFD 670
24 621 NNVNPNLPRTTYKMFSTYRLPVMPELTVGGGVNQONRVYTDVTEYGTFRBAQGSYALVD 680
25 671 AMAAARFPKCLKLOINADNIPFRHRYARVAGANTPNIGSESEKTYWIANLRYSF 722
26 681 LFTROVAVKNPSLQGNVNNLFDKTYDINVEGSIY---GTPNPFISITGTQOP 729

```

C:Genetics:
A:Gene: ECs1480

Query Match	19.58; Score 736.5; DB 2; Length 729;
-------------	---------------------------------------

Matches 211; Conservative 148; Mismatches 308; Indels 105; Gaps 26;

```

QY      3 QENSVFRINNTAATV-LAALSSSVFAAQADLEFTHIKGQSSYNAIYTEKNGDYSSFAVT 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     11 QYQAIKRPILLAGCIMALLPASAFAAPATE-ETVIVEGSAI-----APDGDENDYSVTSTS 66

QY     62 VGTKIPASIREIQSISITNQVKDNRVDTFDQIAKTPELRULSNDGSSVYARGYE 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     67 AGTKQMOTQRDIPQSVYIVSQQRMEDQQLQTGLEWMENTLGISXSQADSDRALYYSRGFQ 126

QY    122 YSEYNDLQ.PAQMQS---INGTL.PNLFAFDRVEYMRQPSGLFSDSGEAGIYVLRKRP 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    127 INNYWDGIPPTYESRWNGLDALSDMALFERVEVVRATGTGMTGNDPSAAILMRKAIAT 186

QY    179 -KAFQHAAGAAGFGTHQYKAEADVSGSLNSDGSVGRVMAQTGVA---SPRAEAKNRHE 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    187 SEEFKQDVAAEYSSNMKERVADIQSPLETDGKIRAI-----VGGYQNDSMLDYNSSEK 242

QY    235 TEYAA-ADWDINBDTVLGAGYL.YQQRHL--APYNGLPA-DANNKL.PSLPHYFVAGDWNK 290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    243 TFFSGIVDADLDGLTLLSAGYEYQRI DVNSPTWGLRPRMTDGSNSYDRASTAPDMAY 302

QY    291 FRKNSHDVFADLKHFGNGGYGVKVMRYSDRDADSNY---AFAGSKLGKMTKP----- 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    303 NDKEIKVFWPTLKQRPADTWQATLNTATHSEVEFDSKNMYVDAVYNNKADGMLVGPSYNGP 362

QY    340 ---AGRGCGNTADDDACAVGLGTB-----IKQKALAFDASYSRPRF--LGNTANEF-- 385
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    363 GPDYVGGTQWNSGKRKYVDALDLFADGSYELFGRQHNLMFGSSYSKONNRYFESSWANI 422

QY    386 VIGADVNRPRSTNQGRRTLLYARGGLALNFRSIPQVDLIANAKQVAGYSHVTANTEND 445
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    423 EIGSFFN-FNG-----NFPQTD-----WSPQSLAQDDTT 450

QY    446 EF-GIYKSKTFPHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKTFYAGAVYVLDNDN 504
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    451 HMKSLYAARVTLADPLHLILGARYTWRVDT----LITYSEKHHTTYAGGLVFPINDNW 506

QY    505 SLYL.SLSQLYTTPQTNLDADGKLKPRQNGQFEVYKGSYMDRDLNARYSFFRMKDKNAA- 563
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    507 SYTYASTTSLFQPNDDSSSKYLAIPITGNNYEELGKSDWMNSKRLTTLTIAIRIEDNVAQ 566

QY    564 ---APLNPNNKKTARYALGRRVMEGVEIEISGAVTPRQMIHAGYS-YLHSQIKTASNSRD 619
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    567 STGTPTPGSNGETARYAAYADGVTSKGYEFELNGAITDMMQLFFGATRYI-----AEDNEG 620

QY    620 DGI.FLIMPRHSANLMTTYQ--VTPELTGGGVNAMSGITSSA-----GHHAGGYATFD 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    621 NAINVPLPRTTYVMFTSYSR.PVMPBELTVGGGVNWQNRHYTDTVTVPYGTFRABQGSYALVD 680

QY    671 AAAAARFTFKLQIQAINDINFNRRHYARVGGANTFNI.PGSRTYTANLRYSF 722

Db    681 LFTFRQVNTNFSLQGNVNMILFDKTYDNTVEGSIYV---GAPRNSITGTGYOF 729

```


A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-729 <STO>
A:Cross-references: GB:AE005174; NID:g12514649; PIDN:AAG5848.1; GSPDB:GN00145; UWGP:Z17
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: fhue
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homology <TNC>
Query Match 19.5%; Score 734.5; DB 2; Length 729;
Best Local Similarity 27.1%; Pred. No. 2.6e-43;
Matches 209; Conservative 146; Mismatches 313; Indels 103; Gaps 24;
QY 3 QPMVSFRINMTAATVLAALSSVFAAQTADLETVHIKQORSYNAIVTEKNGDYSSFAVTV 62
DB 11 QYQAITKPSLLAGCIALALPSSAXAAPATEETVIVEGSAT---APDGDENDYSTVTS 67
QY 63 GTKIPASLREIQSVSIITNQVKORNVDTFDOLARKTPGLRVLSNDDGRSSVYARGVEY 122
DB 68 GTKMQWTRDIPQSVTIVSQRMEDQQLQTLGEVMENTLIGISKQADSRLALYYSRGFQI 127
QY 123 SEYNIIDGLPAQMS---INGTLPNLFAPORVEMRGPGLFDSGEMGIVNLVRKRP- 178
DB 128 DNYWDGIFTYESRWGLDGLSDWALFEVVEVVGATGLMTCTGNPSAIAINWVKHATS 187
QY 179 KAPOGHAAGFTGHQYKAEADVSGSLNSDGSVGRVMAQTYGA---SPRPAEKNRHET 235
DB 188 REFQGDVSAEYSGWKNERYVADLQSPLTEDGKIRARI---VGGYQNDSWLDYRNSEKT 243
QY 236 FYAA-ADWDINDPTVLGAGLYQOORHL--APYNGILPA-DANKLPSLPQHVFGADWKF 291
DB 244 PFGSIVDADLGLTLLSAGYEYQRIQDVNSPTWGLPRMNTDGSNSYRARSAPDWAYN 303
QY 292 KMSNDVFPADLKHYPGNGYGVKVMGRYSDRSADSNY---AFAGSKLGMKTP- 339
DB 304 DKEINKVFTLQKQADTQWATLNATHSEVEFDSKMYVDAYVKNADGMLVCPYNYGPG 363
QY 340 ---AGRPCCNTADDKACAVGLGTE-----IKOKALAFDASYSRPF--LGNTANEF--V 386
DB 364 PDYVGGTGNWGNKRVKVDALDLFADGSEYELFGQHNLMPFGSYSKQNNRYFSSWNIFFDE 423
QY 387 ICADYNRRSTNEQRTTLYARGGLALNEFRSIPQVDLIANARKVGRVSHYTVATENLDE 446
DB 424 IGSFYN-FNG-----NFPQTD-----WSPQSLAQDDTTH 451
QY 447 F-GIYGKSTFHPADGLSLTGGGRGLGHYKIESGEGKTLHKASKTKTPTGVAGAYVDLNDNNS 505
DB 452 MKSLIAARTVTLADPLHLILGARYTNWRVDT---LTYSMKNTHTTPYAGLVFDLNDNWS 507
QY 506 LYLSSLQLYTPQTNLDADGLKLPKQGNQFQYGVKSGYMDRLNARVSFYRMKDKXAA-- 563
DB 508 TVASVTSIFQPNDRDSSKYLAPITGNLYELGLSDWNNSRLTTLTAIFRIEQDNVAQS 567
QY 564 --APLNPNNKTRYAALGRVMEGETEISGAVTPKQIHAGYS-YLHSQIKTASNSRDD 620
DB 568 TGTPTPGSNGETAYKAVDGTGSKGVFEFLNGAITNMQWLTFCATRYI-----AEDNEGN 621
QY 621 GFLMLPKHSANLWTTYO--VTPELTIGGVNMGISGSSA-----GMHAGGYATPDA 671
DB 622 AVNPNLPRTTVMFTSYRLPVNPELITVGGGVNQWNRVYTDTPYGTFRABEGGSVALVDL 681
QY 672 MAAYRFTPKLQIINADNINFRHYVARVGGANTFNIPGSERTWTANLRYSF 722
DB 682 FTRYQVTKNFSLQGNVNNLFDKTYDITNVEGSIVY---GAPRNFISITGYQF 729
RESULT 13
A36942
Fe(III)-pyochelin receptor fptA precursor - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 31-Dec-2000
C:Accession: A36942; G83117
R:Ankenbauer, R.G.; Qian, H.N.

J. Bacteriol. 176, 307-319, 1994
A>Title: fptA, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a phenolate sid
A:Reference number: A36942; MUID:94117363; PMID:8288523
A:Accession: A36942
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-720 <ANK>
A:Cross-references: GB:U03161; NID:g454352; PIDN:AAC43213.1; PID:g454353
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; R
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83117
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-720 <STO>
A:Cross-references: GB:AE004839; GB:AE004091; NID:g9950434; PIDN:AAG07609.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: fptA; PA4221
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom
E:98-227/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:446-720/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
Query Match 18.9%; Score 715.5; DB 2; Length 720;
Best Local Similarity 27.3%; Pred. No. 5.5e-42;
Matches 200; Conservative 137; Mismatches 323; Indels 75; Gaps 20;
QY 21 LSSSVFAAQTAD-----LETVHIKQORSYNAIVTEKNGDYSSFAVTVGTKIPASLRE 72
DB 28 LALSPLAAAVADARKDGETELPDMVISGE---STSATQPPG-----VTLGLKVPKPRE 78
QY 73 IPQSVSIITNQVKORNVDTFDOLARKTPGLRVLSNDDGRSSVYARGVESEYINIDGLPA 132
DB 79 LPQASVTDHERLSEQQQLFSLDEAMQOATGVTVPQFOLLITAYYVRGFKVDSFSLDGVPA 138
QY 133 QMOSINGTLPNLFADRVEMRGPGLFDSGEMGIVNLVRKPTKAFQGHAAAGFQTH 192
DB 139 LLGNTASSPDMAIYERVEILLRSGNLLHGTGNPAATVNLVRKEPREFAFATSTLSAGRW 198
QY 193 KQYKAEADVSGSLNSDGSVGRVMAQTYGASPRP-----AEKNRHETFYAAADWDINPD 247
DB 199 DRYAEADVGVGPLSASGNVRGRAVA---AYEDRDYFYDVADQGTGTR--LLYGVTFBFLDSPD 253
QY 248 TVLGAGLYQQ-RHLAPYNGILPADANNKLPSPQHVFGADWKNFKMNSHVDVADLKHYP 306
DB 254 TLLTVGAQYQHIDSIITNWAGVPMKADGSLNGLSRDLYDWDVDRFKMWTYRAFGLSLEQL 313
QY 307 GNGYGVKGRYSDRDADSNYAFAGSKLGMKTPAGRPCCNTADDKACAVGLGTBIKQKAL 366
DB 314 GGGWKGKVSABY--QEADSRRLRYAGSFGAIDPQTG-----DGGQLMGAAYKFKSQR 363
QY 367 AFDAYSRPFRLGNTANEFVIGADYNRFRSTNEQRTTLYARGGLALNEFRSIPQ--VDL 424
DB 364 SLDLANLGPVRLFLGLTHELLGGVTYAQOQRTDAR-----FLNLPNTPVN 410
QY 425 IANARKGVR---CYSHTVATENLDEFGIYKSTFHPADGLSLTGGGRGLGHYKIESGEGK 480
DB 411 YRWDPHGVPRPQIQYTSYSPGTTTTTQKGLVALGRIKLAEPLTLVVGRESHW--DQDTPA 468
QY 481 TLHKASKTKFTGYAGAYVDLNDNNSLYLSLSQLYTPQTNLDA-DGKLLKPKQGNQFQVGY 539
DB 469 TRFKPGR-QFTPYGGLIWDPAFADWSVYVAEVYQPADRQTNWSEPLSPVEGKTYETGI 527
QY 540 KGSYMDRLNARVSFYRMKDKXAA-----PLNPNKKTRYAALGRVMEGVETEISGAV 594
DB 528 KGEADGLRLNLSLAAFRIDLENNPQEDPHGPPNN--PFYISGGKVRSQGFLEGTGL 585
QY 595 TPQWQIHAGYSYLHSHQ-IKTASNSRDDGIFLLMPKHSANLWTTYQV---TPELTIGGVN 650
DB 586 TPYWSLSAGYTTTSTYELKQSDNSGTRYSTFTPHLLRLLSNLYDLPWQDRRWSVGGGLQ 645

Db 562 SPDYGRSSSTERFQYIGYLOQDTFLDNKLLKLGGRFDMISGENTDNVGTDTQNPDS 621
Qy 498 TKFTGYAGAVYDLNNDNSLYLSLSQLYTPQTNLDAGKLLKPRQNGQPEVGYKGSYMDDR 547
Db 622 SAFPSPRIGLVYQPSKVSLSYTSYSGFVPETGVNPDGGEIFETRTQVEAGIKADFLSGR 681
Qy 548 LNARVSFYRMKDKNAAAFLPNPNKKTRYAAL-----GKRVMEGVETISGAVTPKQIH 601
Db 682 LSATLAAVQITKSNILTP-DPDPER---AALDYLIQVGEQRSGRIELDVAGBILPGWKAI 737
Qy 602 AGYSYLHQSQIKTASNSRDDGIFLL-MPKHSANLWTTQVTP-----LIIGG---VNAMS 653
Db 738 ASYAYTNAEV-TEDNDIPVGNRLVSPKQASLWTTTFQNSDLKGLGFLGLFVYGTGRS 796
Qy 654 GITSSAGMHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGGANTFNIPGSERT 713
Db 797 G-DSANSFEIPDYLKTDAAIYR-RDGFKAGINIRNLFDTDIYRSDGRTFLRGAFTT 854
Qy 714 WTANLRYSF 722
Db 855 IIGSISWEP 863
RESULT 16
AE2130
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2130
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2130
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-820 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB74295.1; PID:gl7131689; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2596
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
Query Match 12.9%; Score 486.5; DB 2; Length 820;
Best Local Similarity 24.6%; Pred. No. 7.1e-26;
Matches 185; Conservative 112; Mismatches 353; Indels 103; Gaps 24;
Qy 15 AFVLAALSSSVFAAQ-----TADLETVHIKQGRS--YNAIVTERKNGDYSSFAVTVGTKIP 67
Db 126 AGLIILAVESTTATQPPATPTSDAPPAEPIAQDDPIELVVTGEQDRVYVTPASTATKTD 185
Qy 68 ASLREIPQSVSITNQVQKDRNVDTFDQLARKTPGLRLVNSDGRSSVYARGYEYSEVNI 127
Db 186 TPVRDIPGSIQVPRQILEQDQKTRIQEVLQNVGVNKGNYGGTD--AGGYRIRGFDQ 242
Qy 128 DGLPAQMSQING-----TLNLFARFVEMRGPSGLFDSGSGMGGINLVLRKPTKA 180
Db 243 DG-----NFRNGFNDTDFYSLVDTANIDRIEVLKGPASVLFQGAEPGGILNVVTKQPLRT 297
Qy 181 FQGHAAAGFTHKQYKAEADVSGSLNSDGSV--RGRVMAQTGASPRPAEKNRHETFYA 238
Db 298 PYAAELNVGNVAFYRPSFDISGLTDDGSLLYRLNVAQNSGSF---RDYNFLERVFA 354
Qy 239 -AADNDINPDTVLGAGLYQQRHLAPYNGLPADANNKLPSPQHPVYFVGADMNKFKQNSHD 297
Db 355 PVITWNISDRTSLTFDLEYODNDYLPDRGIPSGIDRPAF-IPISRFVGLP-HVYNDSTFR 412
Qy 298 VFADLKHPGNGGKGVKNRYSDRADSNYAFAGSKLGMKTPAGPCGTCNTADDKACAVGL 357
Db 413 IGRLEHDFSKDQWLRNAPSFVSGSSGTYAGGYDL-----IDDQFAPITV 459

Qy 358 G-TTIKQKALAFDASYSRPFRLGNTANEFGVIGADYNFRSTNEQGRITTLTARGGLALNEF 416
Db 460 SRDEFTDRDIYTLQTEVVGVKFKTGSIVHQPLIGVELRR-----NTWKYTSFDVADPILLDIF 515
Qy 417 RSIPQVDLIANARKGVGYSHTVATENLDEFGIYKSTFFHPADGLSLIGGRLGHY-KIE 475
Db 516 NPNDVDELPAFPDSESTFSY-----TTRRDITLGIYVQDQITFADNLKULVGGRFDAFORKE 570
Qy 476 SGEKTLHKASKTKFTGTGAGAVYDLNNDNSLYLSLSQLYTPQT--NLDADGKLLKPRQGN 533
Db 571 EGPSETASEESLSAFSPRIGIVQPIQAIISLYASYSQSPKDFRFGRSASNEPKFTRGT 630
Qy 534 QFEVGYKGSYMDRLNARVSPYRMKDKNAAA--PLNPNKKTRYAALGKRVMEGVETEIS 591
Db 631 QYEVGIRKAD-ISEKLSATLAAEYITKTNVVTSDDPNDLS-----VQVGEQRSGRIELDIG 685
Qy 592 GAVTPKQIHAGYSYLHQSQIKTASNSRDDGIFL-----LMPKHSANLWTTQVTPPELTIG 646
Db 686 GEIVPGWNIITASYTY-----TDAITSKDNITPVGNRIDNVNPEHAASLWTSYEL----- 733
Qy 647 GGVNAMSGITSSAGMHAGGYATFDMAAAYRFTPKL-----KLQINADN 689
Db 734 -----QSGDLKGLGFLGYVGDYADVENTSLSSYFRTDSAIYKRONWRLALNFRN 788
Qy 690 IFNRHYARVGGANTFNIPGSERTWTANLRYSF 722
Db 789 LFNETYVETSQARNTI-YPGAPFTVIGSFSIQF 820
RESULT 17
AB1944
ferrichrome iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB1944
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-885 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB73058.1; PID:gl7130447; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all101
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
Query Match 12.9%; Score 486.5; DB 2; Length 885;
Best Local Similarity 23.9%; Pred. No. 7.9e-26;
Matches 179; Conservative 117; Mismatches 343; Indels 109; Gaps 26;
Qy 7 VFRINMTAATYLAALSSSVFAAQTAQTLVTHIKQGRSNAIVTERKNGDYSSFAVTVGTGI 66
Db 191 IFEVAAPTAPTAPQESQVQEQPS-----ASEEPIELVVTGEQGYKVEDATGGTGT 244
Qy 67 PASUREIPQSVSITNQVQKDRNVDTFDQLARKTPGLRLVNSDGRSSVYARGYEYSEYN 126
Db 245 NTPIRDTPFSIQVVPVEEVKDDQVQVTVTDALRAVPG--VISQDAPVSA-----FESFN 295
Qy 127 IDGLPAQMSQINGTLPNLF-----FDRVEVMRGPSGLFDSGSGMGGINLVLRKPT 178
Db 296 FRGSSSGFLNGIRDETIGTAGSGVNIERIEVRGAPALFGSGPGGVNIVTKQPL 355
Qy 179 KAFQGHAAAGFTHKQYKAEADVSGSLNSDGSVGRVMAQTGASPRPAEK-----N 230
Db 356 SSPFVEIGTGVGSDIYEGRVDLTGPLNRDND---TLLYRLTAS---ASKLGGFIDFVD 408
Qy 231 NRHETFYAAADW---DINPDTVLGAGLYQQRHLAPYNGLPADANNKLPSPQHPVYFVGADW 288

```

QY 187 AGFGHXYQKAEADVSSLSNDGSVRGV--MAQIVGASPRAEKXNHETFYAA--ADMD 243
Db 334 MTISGYDFYRTIDTFGTPLTDKKSIVLYRLNVAYNSSGSRFDPIEN--ESFFIPVYTVN 390
QY 244 INPDVLGAGYLYQQRHLAPYNGLPADANNKLPSELQHVHVG--ADWNKFKXNNSHVDFAIL 302
Db 391 ISPTKSMTFEYEQKYNTVFDRL--PGNTFFQIPISRFJGEBGFNNAAEISNVFTYTL 448
QY 303 KHYGNG-----GKYXGMRYSDDADDSNYAFASKL--GMKTPAGRGCTATDDKACA 354
Db 449 EHQSDMKFQRQGNVTSIRGNTRIA--RNTNISEPFLDDQTLPR--TSETDEREEN 504
QY 355 VGLGTEIKOKALADASYSRPFRLGNTANEFVIGADYRFRSTNEOGRTTLVYAGGLALN 414
Db 505 ISLQTEVSGK-----FNTGSI RHNVLLGVELAKKYFT-----Y 537
QY 415 EFRSLP--QVDL---IANARKGV--RGYSHTVATENIDEGITYGKSTPHPADGLSIGG 467
Db 538 DFFAPFASIDIFNPVYGAOPGTFRSFPAGEYGDNL---AVYEQNLIEPTPNKLALAG 594
QY 468 R---LGHYKISGEGKTLHKASKTFKFGYAGAVVDLNDNNSLYLSLSOLYPO--TNLDAD 523
Db 595 RFDMDISDRDPRVSNVTYANNEVSESNFSPRGVIVQPNNSTLSYWNSSWNPQFPGRSRT 654
QY 524 GKLLKPRQGNQFEVYKGSYMDRLNARVSFYRMKDKNAAAPLNPNNKKTGYAALGKRW 583
Db 655 GESFKPETSSEQFEVGIKQEFPRDKRLSATLAFYDTTKNNVLLPPDVDVNNFS--VQIGEGKS 712
QY 584 EGVETETISGATLPKMQIHAGSYLSHSQIKTAS-----NSRDGIFLMPKHSANLMTTYQ 638
Db 713 RGLENDIAGELLPWKKIATAYATYIDSSVSXKONDLERLNDRLISGV---PFSASLMTTYE 768
QY 639 VTPELLTGGGVNANSGITSSAGM-----HAGYAFDAMAAARFPPKTLQ 684

```

Cy 685 INADNIENRHHYARVGAGANTFNIGSERTWTANIRYSF 722
 ||| :|| : : : | : : :
Db 820 INIKNLFDTEYYE---SQSFYLVPAPLTLVTGISFEF 854

RESULT 19

AD2421

ferrichrome iron receptor [imported] - Nostoc sp. (strain PCC 7120)

C/Species: Nostoc sp.

A/Nostc: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C/Acession: AD2421

R/Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

NNA Res. 8, 205-213, 2001

A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A/Reference number: AB1807; MUID:21595285; PMID:11759640

A/Acession: AD2421

A/Status: preliminary

A/Molecule type: DNA

```
A;Cross-references: GB:BA0000019; PIDN:BAH76623.1; PID:g1j134062; GSPDB:GN00119  
A;Experimental source: strain FCC 7120  
C;Genetics:  
A/Gene: all4924  
C/Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
```

```
Query Match          12.7%; Score 481; DB 2; Length 858;  
Best Local Similarity 25.2%; Pred. No. 1.se-25;  
Matches 189; Conservative 127; Mismatches 339; Indels 96; Gaps 27;
```

```
Oy      17 VLAALSSVFPAACADDELTAETHIKQ-----RSYNAIVTEKGDSGFVATGTRIPASL    70  
       ::::|::|::|::||::|::|::|::|::|::|::|::|::|::|::|::|  
Db     159 LIVAVTSAKVMQP--ETPAIPECPAAQDPDEPTELVTGTGGDQRVPAPASVGRTDTPL   216  
  
Oy     71 REIFSVSIIITNQCKDRNVNTFDQLARKTPGLR-VLSNDGRSSVVARGVESEYN---- 127
```

Db 217 RDIPQIQVPPQVQLQEQKATLGLDALRNVSGVNPTRGSGDRADSFTRGFEIFSGNVLN 276
Qy 128 DGLPAQMSINGTLP---NLFAFDRVEVMRSGSLFDSSGEMGGIVNLVRKPTKRAFOGH 184
Db 277 NGLP-----DRTLTRDLYNVERVEVLKGPASVLYGLGNPGGVNIVTKQPLANPFYN 330
Qy 185 AAAGFGTHKQYKAEDVSGSLSDSGSRGRVMAQTVGASPRPAEKNRHETPYAANDWI 244
Db 331 IEATVGNVLYRGGIDFSGPLNDSKTLRLNLAQNQSGSYIDFVGNRSFFTIAPVISAAL 390
Qy 245 NPDVTLGAGLYQORHLAPYN--GLPADANNKLPS-----LPQHVFGADWNKFKWNSHD 297
Db 391 GKNNTLTTEGEYSQKTIIDSTRTVVVLPV--VGTVLPGDGRRIIPNRTVTEPEGDTQIETTR 449
Qy 298 VFADLKHFGNGGYGKVMRYS--DRDADSNYAFAGSKLGM---KTPAGRPGCNTADDDKAC 353
Db 450 LGYRLHFRSENNLSLRNDRFVTFEHNADNNQAF---LGLDADNETANRSTYSSES--SN 505
Qy 354 AVGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNNRFRSTNEQRTTLTYARGGLAL 413
Db 506 IYNLTADI-----SGRFFTGSIHQLLFGVNMRSRFDNF-----NFGIDL 545
Qy 414 NEFRSIPQVDLIANARKGVYSGHVTATEN---LDEFGIYKSTHFPADGLSLGCGRLG 470
Db 546 ABELTFLDIYNPV--YRQPIVGRIDAVIEDNSRLTDTLGIYIQQDKFAENFOLLGGRPD 603
Qy 471 HY-----KIESGEGKTLHKASKTKFTGVAGAVYDLNDNNSLYLSLSQLYTQT 518
Db 604 LFTQKNKFLDNTTELOAGDA-----FTPRGVIVYKPISAISLYASYSQSFPNTE 653
Qy 519 NLDAQGLKLKPRQNGQFVGKYSYMDRLNARVSFYRMKDKNAAAPLNPNKKNTRYA-A 577
Db 654 GRSADGNLFPQPERGTQYEVGVKAD--LNDRISSLTSLYRLTRSNLLT--TDPNN--SRFSIQ 709
Qy 578 LKRVMEGVETEISGAVTPKQIHAGYSYLHSHQIKTASNRDDGIFLLMPKHSANLWTTY 637
Db 710 TGEQRSQGIIEFDIAGEILPGWKIFAGAYTDAEIVEDNTYTSNGRLTNVPEHSFNWTTY 769
Qy 638 QVTPELTITGGVYV--AMSGITSSAG-----MHAGGVATFDAMAAYRFTPKLKLQINADNIF 691
Db 770 EISSGDFRGLGFLGLFYIGDRAGDLNDSFEVPSYLRDTASIFYR--RDLRLPALNKNKIF 828
Qy 692 NRHYARVGGANTFNPISGERTWTANLRYSP 722
Db 829 DTDYFVSVNTRD--FVLRGDPFTISGTISWDF 858

RESULT 20
TonB-dependent receptor [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 14-Sep-2001
C:Accession: D87252
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87252
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-851 <STO>
A:Cross-references: GB:AE005673; NID:gl3421118; PIDN:AAK22016.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0028
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 12.7%; Score 479; DB 2; Length 851;
Best Local Similarity 24.7%; Pred. No. 2.5e-25;
Matches 198; Conservative 126; Mismatches 341; Indels 138; Gaps 33;

Qy 10 INMTAATVLAALSSSVFAAQTADLETVHIKQORSYNALVTEKNGDYSF-ATTV----- 62
Db 95 LSIAAVAGVAGLG---LAQAA-----VAGPDSADAATVTANADDSVSVSTIDAKRVA 144
Qy 63 ---GTKIPASLUREIQSVSIITNQOVKDRNVDTFDQLARKTPGLRVLNSDDGR---SSVY 116
Db 145 DPSSSKTAPLDVTPKSVTVIPAKIIETQATSLADILRTSPGITFGAGGEGQPLADRP 204
Qy 117 ARGVEYSEYN--IDGLPAQMSINGTILNLPADFARVEVMRGPGLFDSSGEMGGIVNLVR 174
Db 205 IRG-QASANNVFDG---VRDSGGQVREIFNLQEVQVVKVGPDSAYGGRSGGGGSLSS 259
Qy 175 KPPTKAFQGHAAAGFTGHQYKAEADVSGSLSDSGSRGRVMAQTVGASPRPAEKNRH 233
Db 260 KSPKADSFARGSVGVGTDAYVVRATADLNHALNESAVRLNLLA--TQGDTPGRKSVSFD 317
Qy 234 ETPVAAADWDINP-----DVLGAGLYQORHLAPYNGLPADANNKL------SL 278
Db 318 -----WGVAPSLATGLDGTQLTATSYHLEGDQTPDYGVPLLTQTPRTASGILDV 369
Qy 279 PQHVFGADWNKFKWNGSHDV--FADLKHYPGNGGYGKVMRYS-----DRDADSNYA 328
Db 370 DRRSFYGVASRDYQKTKSDIATPA-IDHRIDETLNLQVVRYSKSLNDYIVTNPFGDGA 428
Qy 329 -FAGSKLGMKTPAGRPGCNTADDKACAVGLGTEI--KQKALAFDASYSRPFRLGNTANE 385
Db 429 QFVGQGMWKR-----GTKTRWNPETVAAVTDLHGKKTFLGLESHSPDVGLELSREEN-- 481
Qy 386 VIGADYNFRSTNEQ-----GRITLVARGGLALNEFRSIPQVDL---IANARKGVRGY 435
Db 482 -LNATYSTFTTSGAACCTGFTIAATTLASLAGDCTLVYKPNKDQAWTGVINRAPARNV 540
Qy 436 SHTVATENLDEFYIGKSTPHPADGLSLIGGRLGHYKIESGEGKTLH----- 483
Db 541 AKTTA-----LYGPDTVKFGKVLNGLRHRDRESKVDVATTQANGVFTSVTVTP 592
Qy 484 KASKTKTGY-AGAVYDLNDNNSLYLSLSQLYTP-----QTNLDADGK-----LLKP 529
Db 593 RSGSWAFTNYQVGLVYKPTPGSSLYSVYSTASTPPGISAGDQNSNTATGTGNLATVQL 652
Qy 530 RQNGQFVGKYSYMDRLNARVSFYRMKDKNAAAPLNPNKKNTRYAALGKRVMEGVETE 589
Db 653 EDSSEFAGAKANVPHDTLALSALFQTSRKNAQIQID----ATTYAQVGEVKGFEFG 708
Qy 590 ISGAVTPKQIHCAGYSYLHSHQ-IKTASNRDDGIFLL-MPKHSANLWTTVQVTPELTIG 647
Db 709 VSGNITPKQVFGGYTYMDSSELVREGAVTSVNOGDPPLANTPKHSISSFTYKVRKIALG 768
Qy 648 GV-----NAMSGITSSAG-----MHAGGVATFDAMAAYRFTPKLKLQINADNIFNRHYAR 698
Db 769 GAYHVSFKSGNGQGGAGSRIYAPAYWRYDAFASWAVSTGVDLQNLQNLTDERYIAR 828
Qy 699 VGGANTFN-IPGSERTWTANLRY 720
Db 829 TNGVHADPAPGROAILTINVKY 851

RESULT 21
AI2077
ferrichrome iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AI2077
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasanoto, S.; Watanabe, A.; Iriuchi
N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2077
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-863 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA073874.1; PID:g17131266; GSPDB:GN00179
A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alx2175

C:Superfamily: ferrichrome-iron receptor 1; conB-dependent receptor amino-terminal homol

Query Match 12.6%; Score 475; DB 2; Length 863;
Best Local Similarity 25.5%; Pred. No. 4.9e-25;
Matches 176; Conservative 120; Mismatches 301; Indels 94; Gaps 26;

```
QY 47 IYTERKGDVSFAVVTGKIPASLREIPOSVSITNQOVDRNVDFPDQARKTPG-LHY 105
DB 201 VTTGEADSRVPRASATRTDPLRDIPOSIVVPOQVLDKQOVVLTNEMLQVNSGYLTQ 260
QY 106 LSNDDRSSVYARGYEV-----SEYNIDGLPAQMOSINGTLPLFADRYEVRGSPGLF 160
DB 261 ASVYSGFAFTIRGFNSFDQGNPFTRNGLYGRFGS-QGT-NFSNIERLEVLRGSGVYL 317
QY 161 DSSGEMGGIVNLVYRKPTAFQGHAAAGFETHQYKAADVSGSLNSDGSVGRGVMAQTV 220
DB 318 FQSGNGGGINITYKQPLSEPFYSVEAIGSYDFYRGALDLSGPLDDSKTALYR-----L 372
QY 221 GASPRPAEK--NNRHETFYAAH--DMDINPDTVLGAGVLYOQRHLAPYNGLPADANNKL 275
DB 373 NASYEKADNVDFNDRENSVASTLSFALGENTTLTLDQETNKNVNGYINGVPA-VGTVL 431
QY 276 PSL-----PQHVFGADNMKFKMNSHDVFADLKHYPGNGYGVKVMYS--DRDASNYA 328
DB 432 PNLNGRIPIRRRSIGQADSTYSPETIVRVGNVLEHKFSEDMILRNAPFYSHYNNTRDTYFA 491
QY 329 FAGSKLGKMTPRAGPCGNATDDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFGVIG 388
DB 492 NA---LDPDORTLQORVODADDRVQTYDLSINYVGK-----FSTGSIKHOLLFG 537
QY 389 AAYNRFRSTNEOGRTLYARGLALNEFRSIPQVD---LIANARKGVYSTVATENL 444
DB 538 VDISR-----DQYQSEYEGRTGTPDLFENDYRPERFEVFA-----GNEATLT- 582
QY 445 DERGIYKSTFHPADGLSLIGGRGLGHYKIESGE--GKTLHKASKRTFTGYAGAVYDLND 502
DB 583 DSLGVYIQGVFAENPFKLLGSRFPIFEQTKTDRLSNTQFQSGSAFSPRGVIVYQPIR 642
QY 503 NNSLYLSLSQLYTPQTNLDADGKLKPRQGNQPEVGVKGSYMDRLNARVSFRMKDKNA 562
DB 643 PSLVASYRSRFPITIGRASDGEQFPGRGTOYEIGVKAD-INEKLSATLAFDLTRSNV 701
QY 563 AA--PLNPNKKTRVYALGKRWMEGVETELSGAVTPKQIHAGYSYLHSQIKTASNRD 620
DB 702 TTNDPAPNPPS---IQTEQNSRGIELNVAGELPLGMMNIIAGYAYTDARI-TQDNSLPI 756
QY 621 GTFLL-LMPKHSANLMTTYQVTPPELTIGGVNAMSGITSSAGMHAG----- 665
DB 757 GKRLLNVPEHSLSLMTTY-----ELQGN-----LQGLGFLGLFYLDROGDLANSFNLS 808
QY 666 YATFDMAAAYRFTPKLKLOINADNIFNRHY 696
DB 809 YLRDAAIFYK-RDRFRALMINRLFDLKYF 838
```

RESULT 22

E64817

probable membrane protein b0805 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: E64817

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

C:Accession: E64817

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-760 <BLAT>

A:Cross-references: GB:AE000182; GB:U00096; NID:g1787015; PIDN:AACT3892.1; PID:g1787024,

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: ferrichrome-iron receptor 1; conB-dependent receptor amino-terminal homol

C:Keywords: transmembrane protein

F:1-49/Region: conB-box

F:17-33/Domain: transmembrane #status predicted <TM>

Query Match 12.4%; Score 470; DB 2; Length 760;
Best Local Similarity 22.8%; Pred. No. 9e-25;
Matches 170; Conservative 109; Mismatches 331; Indels 134; Gaps 23;

```
QY 65 KIPASIREIPQSVSITNQOVDRNVDFDQARKTPGL-RVLSNDDGS---SVYARG 119
DB 63 KPSRPVADTTMTVISEQVIXDQATNLTDALKVPGVGAFFAGENGSTGDAIYMG 122
QY 120 YEYS-EYINIDGLPAQMOSINGTLPLFADRYEVRGSPGLDSSGEMGGIVNLVYRKPT 178
DB 123 ADTSNSIYIDG---RDIGSVSRDTFNTQYEVVIXKPSGTDYGRSAPTSINMISKOPR 178
QY 179 KAFQGHAAAGFETHQYKAADVSGSLNSDGSVGRGVMAQTVGASPRPAEKNNRHETFYA 238
DB 179 NDSGIDASISGSAMRRGTLQVIGDTTAVRLNVMEKTHDAGRKVKR----- 232
QY 239 AADWDINPDTVLGAG-----YLYQQRHLAPYNGLP-----ADANKLPS 277
DB 233 ---YGVAPSVARGLGTHANLYLVNTHVQHTNPDDGIPITGLPGYSPAGTALNHSK 289
QY 278 LQHVNVGADNMKFKMNSHDVFADLKHYPGNGYGVKVMYS--DRDASNYA 334
DB 290 VQTHNHYGTDSDYDSDTDTATMRREHDINDTTIRNTTMRVRYQD--YLMTALMGAS 347
QY 335 GMITPRAGP---CNTADKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFGVIGAD 390
DB 348 NITQPSDVNSMTWRTANTK-----DVSNKILTNQNTLSTFTYSIGHDSTGVB 399
QY 391 YNRFRSTNEOGRTLYARGGL--ALNEFRSIPQVDLIANARKGVYSTVATENLDEF 447
DB 400 FRETQTN-----YGVVPVTLPAVNIYHPDSSIHPGLTRNGANANGOT-----DTF 446
QY 448 GIYKSTFHPADGLSLIGGRGLGHYKIE-----SEEG-----KTLHK 484
DB 447 AIIYAFDTLQITDFELNGGIRLDNHTFEYDSATPACGSGRAITCPTGVAKSPVTTVOT 506
QY 485 ASKTKFTGY-AGAVYDLNNSLYLSLSQLYTPQ-----TNLPADGKLKPRQ 532
DB 507 AKSGNLMKAKGLHYLTNGVNTINYAVSQPPGANNFALAQSGNSANRTDFKQQA 566
QY 533 NOFEVGVKGSYMDRLNARVSFRMKDKNAAPLNPNKKTRVYALGKRWMEGVETELSG 592
DB 567 NNSLIGTKNQVLDKRLLLPALFRDTIENEV---EQNDGTYSGYQKRVGEISVAG 622
QY 593 AVTPKQIHAGYSYLHSQIKTASNSRBDGIFLL--MPKHSANLMTTYQVTPPELTIGGGIN 650
DB 623 NITPAQVIGVYTOQKATIKNGKDVADQSSSLPYTPBEAFTLMSQYQATIDISVAGAGR 682
QY 651 AMSGITSSAGMHAG-----GYATPDMAAAYRFTPKLKLOINADNIFNRHYAR 698
DB 683 YI-----GSMKSGDGAAGVTPAFTGIVVADAKIGYVRNRLDQNLNLYNLPDDIYAS 736
QY 699 VGGANTFNIPGSERTW--TANLRY 720
DB 737 INKSGYRYRGPFRPFLLANMHF 760
```

RESULT 23

AC2140

ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AC2140

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Saeamato, S.; Watanabe, A.; Iriuguhi

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-867 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074373.1; PID:gl7131767; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2674
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 12.4%; Score 467.5; DB 2; Length 867;
Best Local Similarity 24.7%; Pred. No. 1.6e-24;
Matches 186; Conservative 127; Mismatches 264; Indels 175; Gaps 38;

Qy 61 TVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDQLARKTPGLRVLSNDDGRSSVYA--- 117
Db 201 SVGTRDTPLIDVPOAIOVPOEVINDQGRSLGKNTS-----SASSGRSTSQAPAL 255
Qy 118 ----RGVEYSEYNIDGLPAQMSINGTLNLFAPDRVEMRGPSGLFDSGSGEMGIVNLV 173
Db 256 TPVIRGFESNLLRNGRLDSDLSRFQSEIANV---ERVEVLKGPASVLFGQDGLGGVNLV 312
Qy 174 RKRP--TKAFQ--CHAAAGFTGHQYKAEADVSGSLNSDGSVRGRVMAQTGVGASPPAEK- 229
Db 313 TKQPLNTPYISIGYQVQGLH--RPTIDFSGPLDKDG-----VAYRLNAAQTAESEF 363
Qy 230 --NNRHETPYAADWDI--NPDTVLGAG--YL-YQORHLAPYNGLPA----- 269
Db 364 KDFENSESEFFIAPVVRLLIGNENTNLTAIEYLKYSFETAP--DLPASGTVISNPNGRVS 421
Qy 270 -DANNKLPSLPQHVFGADWNKFKMNSHDVADLKHFGNGGYGVGMRYSDR---DADS 325
Db 422 RETNLGEPSLSE-----SESLVRL-----GY-QLDHLNDNWTIKSEF 459
Qy 326 NYAFAGSKLGMKTPAGR---PGCNTADDRKACAVGL---GTEIKQ-----KALAFDA 370
Db 460 STAF-----LDVPESTVVLPIANSSTSN---GLNRDGRTLRRFLVENPSSSTTSITFNN 509
Qy 371 SYSRPFRLGNTANEFVIGADYNFRSTNEQGRITLYARGGLALNEFRSIPQVDLI--ANA 428
Db 510 SLLGKPKTGSIEHSLLFGEV-----TRETEDRL-----DFRQLRDIDIFNPVYR 555
Qy 429 RKGVRGYSHTVATENLDE--FGIYKSTFHPADGLSLIGGRLG---HYKIESGEGKTL 482
Db 556 PESVSSFAIPFGNTWTEKNSLGIYAQDQISLSKNILVLGGRLDFWNQDYEDLLSEBESF 615
Qy 483 HKASKTKFTGYAGAVYDLNNDNSLYLSLSQLYTP-----QTNLDADGKLLKPRQG 532
Db 616 ER-NDTVESPRGIVYKPSENLSLYASYSRSFTPVVGRTRVLDTNTGITTVEGEPEPERG 674
Qy 533 NQFEVGYKSGMDDRLNARVSYRKKQKAAA-----PLNPNKKTRVAALGRVMEGVE 587
Db 675 TOYEVGLKANLLGDRSLSTLAFNLNERTNVAAGLSEPLS-----QIQIGKORSQIE 727
Qy 588 TETISGAVTPKQWQHAGYSYLHSGQIKTASNR--DDGIFLLMPKXANLWTVQVTPELTIG 646
Db 728 LDVAGEILPGWNLTAASYATDSKITEDSRPEFQDQLQNVPRNSFGLWSTY-----ELQAG 783
Qy 647 GGVNAMSGITSSAGMHAGG-----YATFDAMAAAYRTPKLLQINADNIFN 692
Db 784 ----SLKGLGFLGVFTQGERQGLRLNTFTLPSYLRTDASIFYR-RDKFRAAINIQNLFD 838
Qy 693 RHYIYARVGQANTF--NIPGSERTWTANLRYSF 722
Db 839 ENYIE---GARDIVRVIPGAPFTLTGSVSFEF 867

RESULT 24
S74450

ferrichrome-iron receptor 2 - Synchocystis sp. (strain PCC 6803)
N:Alternate names: protein sll1406
C:Species: Synchocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74450
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74450
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-828 <KAN>
A:Cross-references: EMBL:D90899; GB:AB001339; NID:gl651650; PIDN:BAAL6602.1; PID:gl6516167
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: ihua_2
A:Start codon: GTG
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
C:Keywords: iron transport
F:213-346/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:547-828/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 12.3%; Score 465; DB 2; Length 828;
Best Local Similarity 23.1%; Pred. No. 2.3e-24;
Matches 182; Conservative 119; Mismatches 303; Indels 184; Gaps 30;

Qy 17 VLAALSSVPAA---QTADLETVHIKQORSYNAIVTEKNGDYSSF--AVTVGTIKIPASLR 71
Db 143 VLSSTSTATAIPENPESEIEW-----ATQEQGEASVFPFSASTATGLDTPLL 192
Qy 72 BIPQSVSIITNQOVKDRNVDTFDQLARKTPGLRVLSNDDGR--SSVYARGYSEVNI--- 127
Db 193 DIPQSIQVVPQVQDRNVTELGALQTVPG---VSPAGRGTSVFGGFLIRFPVNS 249
Qy 128 ---DGLPAQ--MOSINGTLPNLFADPVEVMRGPSGLFDSGEMGIVNLVRKRTKAFQG 183
Db 250 IFRDGIPIQSLAPLNTT-----DIEQIEVLKGPSSIVFGAGEPGGSINLISKPLDEPY 304
Qy 184 HAAAGFTGHQYKAEADVSG-----SLNSDGSVRGRVMAQTGVGASPPA 227
Db 305 NAAVSLGNYNDYRLDVLSCGPLPEAIDTVNRLNVSYETSGSFRDFYIGDLWVSP--- 361
Qy 228 EKNRHETFYAAADWDINPDTVLGAGLYQORHLAPYNGLPADANNKLPSLPQHVFVGAD 287
Db 362 -----TLTNIGPDTKLNLYQYTNRTILDEGIPAP---NIADLPNRFLGER 407
Qy 288 WNKPKMNSHDVADLKHFGNGGYGVGMRYSDRADSNTAFAGSKLGMKTPAGRPGCNT 347
Db 408 FSKFEQDQYLYGTFTNDFNENLKLHAMQY-----LAYAPRYA-----PLPDF 452
Qy 348 ADDKACAVGLGTEIKQALAFDASYSRPFRLGNTANEF-----VIGADYNFRSTN 398
Db 453 FDEPT-----GELNRFPEYGGGNYQRFFTNAELIGEFTYGPVXRVHLFGLYEYRNDTETP 506
Qy 399 EQGRTTLYARGGLALNEFRSIPQVDLIANARKGVRGYSHT---VATE-----NLDEFIY 450
Db 507 EFQFSNTFA---PINVFNPV-----YTNTPFPIAPEFFRDQDQNRFAVY 547
Qy 451 GKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKT-----DSATQNRSTQSIITDPREEFNQTDNQLTPRVGII 497
Db 548 LDQWMDLFDNLKLLVGLRY-----DSATQNRSTQSIITDPREEFNQTDNQLTPRVGII 599
Qy 498 YDLNNDNSLYLSLSQLYTPQ--TNLDADGKLLKPRQNGQFEVGYKSGMDDRLNARVSFY 555
Db 600 YQPIPTVSLGYSYTTSPNPSFAASLNADGSTFDQTRQRFQEVGVKAD--ITDKLSVTFSAF 658
Qy 556 RMKDKNAAAPLNPNKKTRVAALGRVMEGVETEISGAVTPKQWQHAGYSYLHSGQIKTAS 615
Db 659 DIRKQNVPT-IDPANLLFTIQT-GEQTSRGVELYLGLGELLPGWNLTVGYSYLDAPVSDQN 716

Oy	616	NSRDGIEILWPKSHANLMTTYQVYVPELTIIGGVNAMSGITSSAGMHG-----	664
Db	717	TDIVNTLSNVEPSNFSIMTYEI-----QSGNLOGLQFGLPYVDQREGDL	764
Oy	665	-----GATPEDAWAARYFTPKLQINADNIENRHYYARVGANTFNI---PGSERTW	714
Db	765	DNTFVLPSTFRTDAIAYR-RENMELQINIENTFYQYLAE---SNDPDLVYPGAPFTV	820
Oy	715	TANLERSF	722
Db	821	VKGIGVTF	828

RESULT 25
AE2129
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2129
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaamoto, S.; Watanabe, A.; Iriuguchi,
N.; Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2129
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-851 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074287.1; PID:g17131681; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al12588
A:Superfamily: ferrichrome-iron receptor 1, tonB-dependent receptor amino-terminal homol

Query Match	12.2%	Score 460.5	DB 2	Length 851
Best Local Similarity	23.6%	Pred. No. 4.9e-24		
Matches 173	Conservative 120	Mismatches 316	Indels 123	Gaps 25

[illegible]

```

Db      594  DFTNGSPSIDYDVGFKFDISDQAFSPRIGIYYQPTPLSLVASTSRSPAQFGIGRDNMR 653
QY      527  LKPRGNOFENYGYKSSVYDDRLNARVSFYRKMDKRAAP-----LNPNKKTRRYAALGRV 582
Db      654  IEPGEQTQIEGVKREELFDGKLIASLAGTQITKTVAIPDPADLNFS-----IPVGEVR 707
QY      583  MEGVETEISGAVTPKWOQIHAGSYSLHSQIKTASNSRDDGIFL-IMPKHSANLMTWYYQVTP 641
Db      708  SRGIETFDIAGELAKOMNIASAYATDAKI-TEQTDNDEGRNLNVPEPENSASIMWTTEYL-- 764
QY      642  ELTIGGVANASGITSAGMHAG-----GYATPDMAYRYFTPKLKL 683
Db      765  -----QSGALQIGMGVGLFFVEGERQGLSNSFTFYGYRTDPALELYR-RDMNNI 813
QY      684  QINADNITNRKH 695
Db      814  GINFNKIPDVNI 825

```

RESULT 26
 C90739
 Hypothetical protein ECs0883 [imported] - *Escherichia coli* (strain O157:H7, substrain R1)
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001
 C:Accession: C90739
 R:Hayashi, T., Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gatsawara, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genetic
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90739
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-760 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA834306.1; PID:G13360342; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 C:Gene: ECs0883
 C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match	12.2%	Score 460;	DB 2;	Length 760;
Best Local Similarity	22.6%	Pred. No. 4.5e-24;		
Matches 168;	Conservative 109;	Mismatches 333;	Indels 134;	Gaps 23;

```
QY      KIPALREIPOSVSIITNQVKDRNVDFEDQLAKTEGL-RVLNDDORS-----SYARG    119
Db      63 KFSPRPAOTTRMTVVISQVIKIDOGATNLTDALKNPVPVGAFPGEGENSNSTTGDAITYRG    122
QY      120 YEVS-EVVIDGLPAMOMINGTLPLFAFDRVEVMRGSGSFPDSSGEMGIIVNLVRKRPT    178
Db      123 ADTSNSIITIIGT---RIDGSVRPDTEFOEVAIKGPSGDIYGRSAFTGSINISIKOPR    178
QY      179 KAFQCHAAAGCFTHKOYAEADVSGLSNDSYSGRVAQTGVASPPRAEKNNRHEFYA    238
Db      179 NDSGISDABASIGSANFRRCGTLDVNVOIGDTAYAVLINWGEKTHBAGRKYNER-----    232
QY      239 AADWDINDPYLAG-----YLXOQRHLAPYNGLP-----ADANKLPS    277
Db      233 ---YGVAISIAPFGJLTANRLYLIHTYQNTPPGGIPTTGLPGVASAPSACTATILNSGK    289
QY      278 LPQHVFEVGAOWNKEFKMNSHDVFADIKAHYFGNGGYGKQGMRYSDRDADSNY--APAGSKL    334
Db      280 VDHNFPEYSTDSDPYDSTDTJATMRFEHDIINDNTIIRNTFRSRVQOD--YLMTAIMGAS    347
QY      335 GMKKTAGRPG-----CNTADDACAVALGETELKOKALAFDASYSPFRHGIANAFVIGAD    390
Db      348 NITQTPSDVNSWTWSRTINTK-----DVSNKITLNQUTLTSTFYASIGHSVTGEV    399
QY      391 YNRPFSTMEQGRTLLYARGGL---ALNEFRSIPQVDLIANARKKVORGSHVATENDIEF    447
Db      400 FTREFQTN-----YGVNPLYTLPAVNIYHPDSSIHPGGLTRNGANANGQT-----DTF    446
```


Qy 448 GIYKSTFHPADGLSLIGGRLGHYKIE-----SGEG-----KTLHK 484
 Db 447 AIYAFDTLQITRDFELNGGIRLDNTHTEYDSATACGSGRGAIITCPAGVAKGSPVTTVD 506
 Qy 485 ASKTKFTGY-AGAVYDLNDNNSLYLSQLYTPQ-----TNLDADGKLLKPRQG 532
 Db 507 AKSGNLVNWKAGALYHLTENGNNVINYAVSQPPGCGNNFALAQSGSGNSARTDFKPKA 566
 Qy 533 NOFEVGYKSGVMDRLNARVSFYRMKDKNAAAPLNPNKKTRYAALGRVMGVEVETISG 592
 Db 567 NTSEIGTKWQVLDKELLTAALFRDITENEV-----EQNDDGTYSYQKKRVEGEISVAG 622
 Qy 593 AVTPKQIHAHGYLSHLSQIKTASNRDDGIFLL--MPKHSANLWTTTYQVTELTIGGVN 650
 Db 683 YI-----GSMHKGSDGAVGTPAFTEGYWVADAKLGYVRNRLDFQLNVLNLFDTDIYAS 736
 Qy 699 VGGANTFNIPGSERTW--TANLRY 720
 Db 737 INKSGYRHPGEPRTFLLTANMHF 760

RESULT 27
 E85589
 hypothetical protein Z1026 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: E85589
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: E85589
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-760 <STO>
 A:Cross-references: GB:AE005174; NID:gl2513801; PIDN:AAG55177.1; GSPDB:GN00145; UWGP:Z10
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z1026
 C:Supfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 12.2%; Score 460; DB 2; Length 760;
 Best Local Similarity 22.6%; Pred. No. 4.5e-24;
 Matches 168; Conservative 109; Mismatches 333; Indels 134; Gaps 23;

Qy 65 KIPASLRIPQSVSITNQVKDRNVDTFDQLARKTPGL-RVLSNDDGRS----SVYARG 119
 Db 63 KFSRPVADTRTWTWVISEQVIKQCATNLTDALKNVPGVGAFFAGENGNSITGDAIYMRG 122
 Qy 120 YEYS-EYINIDGLPAQMSINGTLPLNFPADRVEMVRGSGSLFDSSGEMGGIVNLVRKAPT 178
 Db 123 ADTNSIVIDGI-----RDIGSVSRDTFTNQVEVIGKSPGSDTYGRSAPTGSINMISKQPR 178
 Qy 179 KAFQGHAAAGFTQKQYKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEKNRHETFYA 238
 Db 179 NDSGIDASASIGSAWFRRTGLDVNQVIGTTAVRLNVWGEKTHDAGROKVRNER----- 232
 Qy 239 AADWDINPDTVLGAG-----YLYQORHLAPYNGLP-----ADANNKLPS 277
 Db 233 ---YGVAPSIAGLGTANLXLYNLHVHTQNTPDGIGTIGLPGVSAPSAGTATLNHSGK 289
 Qy 278 LPQHVFGADWNKFKNSHDVFDADLKHFVNGGYGKVMRYSDRADSNY---AFAGSKL 334
 Db 290 VDTNHYFGTSDSYDDSDTDTATMRFEHDINDNTTIRNTRWSRVKQD---YLMTAIMGAS 347
 Qy 335 GSKTPAGRPG-----CNTADDKACAVGLGTEIKQKALAFDASVSRPRLGNTANEVIGAD 390
 Db 348 NITQPTSDVNSWTWERTANTK-----DVSNKILTNQNTLTSTFTYASIGHDVSTGVE 399

Qy 391 YNPRSTNEQGRITTLIYARGGL---ALNEFRSIPQVDLIANARKVGRVGSHTVATENLDEF 447
 Db 400 FTRETQTN-----YGVNPFVTLPAVNIYHPDSSIHPGGLTRNGANANGQT-----DTF 446
 Qy 448 GIYKSTFHPADGLSLIGGRLGHYKIE-----SGEG-----KTLHK 484
 Db 447 AIYAFDTLQITRDFELNGGIRLDNTHTEYDSATACGSGRGAIITCPAGVAKGSPVTTVD 506
 Qy 485 ASKTKFTGY-AGAVYDLNDNNSLYLSQLYTPQ-----TNLDADGKLLKPRQG 532
 Db 507 AKSGNLVNWKAGALYHLTENGNNVINYAVSQPPGCGNNFALAQSGSGNSARTDFKPKA 566
 Qy 533 NOFEVGYKSGVMDRLNARVSFYRMKDKNAAAPLNPNKKTRYAALGRVMGVEVETISG 592
 Db 567 NTSEIGTKWQVLDKELLTAALFRDITENEV-----EQNDDGTYSYQKKRVEGEISVAG 622
 Qy 593 AVTPKQIHAHGYLSHLSQIKTASNRDDGIFLL--MPKHSANLWTTTYQVTELTIGGVN 650
 Db 623 NITPAWQVIGGYTQOKATIKNGKDVADQDGSLSLPTPEHAFTLWSQYQATDDISVGAGAR 682
 Qy 651 AMSGITSSAGMHAG-----GYATFDMAAAYRFTPKLKLQINADNIFNRHYIYAR 698
 Db 683 YI-----GSMHKGSDGAVGTPAFTEGYWVADAKLGYVRNRLDFQLNVLNLFDTDIYAS 736
 Qy 699 VGGANTFNIPGSERTW--TANLRY 720
 Db 737 INKSGYRHPGEPRTFLLTANMHF 760

RESULT 28

A83043
 probable outer membrane protein PA4837 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: A83043
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.B.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: A83043
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-708 <STO>
 A:Cross-references: GB:AE004896; GB:AE004091; NID:g9951094; PIDN:AAG08222.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4837

Query Match 12.2%; Score 459; DB 2; Length 708;
 Best Local Similarity 22.7%; Pred. No. 4.8e-24;
 Matches 172; Conservative 135; Mismatches 329; Indels 122; Gaps 25;

Qy 22 SSSVFAAQT---ADLE--TVHIKQORSYNAIVTERKNGDYSSFAVTVGKIPASLRIPQS 76
 Db 16 SSPALWAETESPALEVLTVTAEEAEGPV-----QYRANRTASATRTDIEDIPQA 70
 Qy 77 VSIITNQVKDRNVDTFDQLARKTPGLRVLSNDDGRS--SVYARGYSEYNIDGLPAQM 134
 Db 71 ISVVPRLVLDLSDSARIERALDFAGGVSQRNNFGLTFEYNVRGFTTSSEYRQFGFSANR 130
 Qy 135 QSINGTLPLNFPADRVEMVRGSGSLFDSSGEMGGIVNLVRKRPTKAFQGHAAAGFTGHQ 194
 Db 131 GYMNA--PDSATIERVELKGPASSLYCRGDPGCTVNLVTKKPOAERPARLHASAGSWDR 188
 Qy 195 YKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEKNRHETFYA-----AADWDIN 245
 Db 189 YRSTLDLNTPLDEEGDLLYRM-----NLAVEDSKGFRDYADGQRLLVAPSISWOLD 239
 Qy 246 PDT-----VLGAGYLYQORHLAPYNGLPADANNKLPSLPQHVFG--ADNKKFQMNSHD 297


```

Dh 240 PDSLLVAEVYVRNRQVDFRGTVAPH-----NHLSLPSRFRFFGEEDOCKIDNNBT 291
Qy 298 VFADLKHIFGNGGIVKVGKRSYSDRADSNYAFAGSKLGKMTKPPAGPCGCTADDAKCAVGL 357
Db 292 LQATLRHHFNEQWMSLRLASHYKHGHLD-GYASENSLSLA-----ADGYSL 334
Qy 358 GTEIKOKLAAPDASYR-----PFLIGNTANEFVIGADYNRPRSTNEOQRTTLVARGGLA 412
Db 335 RREYRYRPFEMHDSITQDLDLGDHTGSLRQGLMGLERYKH--NDE-----LI 382
Qy 413 LNEFSLIQOVDLIA-MARKVARGY-----SHYVATENLDEFGIYKSTFHPADGL 461
Db 383 L--RSISPRNPYAIIDIRRPVYGQKPPFGRRDRNH-----EEVDAMALNLOQIIEFSKM 435
Qy 462 SLIGGGRIGHY-----KIESGE--GKTLHKASKTKFTGYAGAAVVDLDLNNLSLYLSLQ 513
Db 436 RGLLGVREDRTRQODNNATRLNNGRFRFETSSQOTQAAITPRIGVILQATPEVGLFANASIS 495
Qy 514 YTPQTNLDADGKTLKPROGNOFEVGYKGSYMDRLNARVSFYRMKDKN--AAAPLNPNNK 571
Db 496 FKPNQGTBMAGKAPFPEBGRGYEAGVKLLDLGDRGLMTLAAPHLLKKQVLTADPNSPGVQ 555
Qy 572 KTRVAALEKRWVEGJETELSGAVTPKMQIHNQYSVLSHQIKTASNSRDDGIFILMPKISA 631
Db 556 QT---APABNSQGFDLQSGQLTEQLRLIGAAYAITDDEVTDENRIAGSRLLNPKSG 611
Qy 632 NLMTTYQVTPBELTIGGVVA-----MSGITSSAGMAGGATPDAMAAYFTPEKLLQ 684
Db 612 SLMGVYEPREBGLHGDADGAANVYVGERAGDSSDSGFELPATYTDLLARYLAAENATLIG 671
Qy 685 INADNIENRHHYARVGCANTENIPGSEBRTWTANLRYSF 722
Db 708 VVNNVLFDRRYER--SYNNVWVAPEPRULTMISLTNLY 708

```

```

RESULT 29
G97408
fega protein U61401 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C|Species: Agrobacterium tumefaciens
C|Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C|Accession: G97408
R|Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.,
Science 294, 2232-2328, 2001
C|Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A|Reference number: A97359; PMID:11743194
A|Accession: G97408
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-747 <KUR>
A|Cross-references: GB:AE007869; PIDN:AAK6224.1; PID:G15155326; GSPDB:GN00169
C|Genetics:
A|Gene: AGR_C_718
A|Map position: circular chromosome

Query Match          11.9%; Score 450.5; DB 2; Length 747;
Best Local Similarity 25.0%; Pred. No. 2e-23;
Matches 195; Conservative 104; Mismatches 358; Indels 123; Gaps 33;

Oy      2  GGFMSVFPIINMTAAVTLALSSVFAAQTADLETHIKGGRVYNAIVKXKNDYSSFAVT 61
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      24  GNFELRLALGTRAAVVLMPGVMG--HAQETTVLKQITVEGQGENA--TGPVRGYVAKKSA 79
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy      62  VGTKIPASLRREIPQSVSIITNQVXDRN-VDTFDOLAKRTPELRY--LSNDGRSSVYAR 118
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      80  TGSKTETETKALPGQSVSVVGRQEMDRKAVTKIDELVLRTPETVLAEPRTDPTDITWPFIR 139
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy     119  GYEYSEYNI--DGLPAQMSINGTLPNLFAFDREVMVMPGSLGFDSSGMEGIYALVRRK 176
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     140  GQOATGTGTFDLGLNLFYSYFGFGQMDAYGLERVAEVLKGPASVLYGKANPGGIVQWVSKR 199
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy     177  ----PTKAFQ-----GHAAAGF-----GTHKQYKAEADVSGSLN-SLGS--VGRG 214
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

[illegible]

```

RESULT 30
A13197
TonB-dependent receptor [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid
C1:Species: Agrobacterium tumefaciens
C1:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
R1:Accession: A13197
R1:Mood, D.W.; Seibubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, F.; Zhang, S.
Science 294, 2317-2333, 2001
A1:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A1>Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A1:Reference number: AB2577; PMID:11743193
A1:Accession: A13197
A1>Status: preliminary
A1:Molecule type: DNA
A1:Residues: 1-819 <KOR>
A1:Cross-references: GB:AE006887; PIDN:AAL45999.1; PID:g17743754; GSPDB:GN00188
A1:Experimental source: strain C58 (Dupont)
C1:Genetics:
A1:Gene: Atu5311
A1:Genome: plasmid

Query March 11.9%; Score 450; DB 2; Length 819;
Best Local Similarity 25.5%; Pred.No.2.5e-23;
Matches 187; Conservative 106; Mismatches 349; Indels 92; Gaps 31;

QY 33 LETYHKGGRSYNAIVIEKNGDVSSFAVTGTRKIPASIRREIPOSVSIIITNOQVDNRVDT 92
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 134 LGTTTGGAGSAAHYGV--DGIVASRSAT-GTKIDPTPLIEVPOGISVITLDEVAARGAET 189
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 93 FDQLARKTPGLRVLSNDNGRSSYFARGYEYSVIIDGLPROMOSINGT-LPNL----- 144
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

Db 190 IKEAVNYTAGVHV-----GGSSASTRAFDNIE--IRGFAPTLPLDGLTYPYIGDLGGSP 242
Qy 145 ----FAFDRVEMRGPSGLFDSSGEMGIVNLVRKRPYKAFQGHAAAGFGTHQYKABAD 200
Db 243 QIDPYLELIEVLKGPSSVLYQNYPGGMNVMKRPDKPPNEVVAGTGT-----DGR 296
Qy 201 VGSLSNDSQSVGR--VMAQTGASPRPAEK--NNRHETFYAAADWDNDPVLGAGYL- 255
Db 297 AYGAFFDFSGPVAGNDAFLRLTGVAITRTGNTIDYTKDERFMLAPSFALKPDBDITFTTFLS 356
Qy 256 -YQORHLAP-YNGLPADANNK--LPSLPOHVFVGAD-WNKFKMNSHVDVADLKHVFGNG 309
Db 357 HYKONGVPDYQPLPIGVTVKAGPVGAINRDEFTTEBPAYNGVDRQAVLGVYEFKQFDDV 416
Qy 310 GYGKVGMYSDRDADSNYAFAGS--KLGKMKTPAGPGCNTADDKACAVGLGTEIKOKALA 367
Db 417 WSIIRNAKIYISVDDSVRTFSGGYVETGGVTDYTKMRRNAIDYSS-----NNQVFA 467
Qy 368 FDASYRPPRLGNTANEFVIGADYNRFRSTNEQRTTLYARGGLALNFRSIPQVDLIAN 427
Db 468 TDTNLQAEFTSTGEVGHVTVGTGDKWFRN-DYTGR---YGFNGTEPLDVFN--PRYGSYRE 521
Qy 428 ARKGVRGYSHTVATENLDEFGLYGKSTFHPADGLSLIGGRLGHYKIESGE---GKTLHK 484
Db 522 PTVGARW-----DNRIQLGLYAQDQI-KWDNWLITLGGRY-DNALQTDNDILSTIGK 573
Qy 485 ASKTRFTGAGAVYDLDNNSLYLSLSQLYTPQTNLADGKLLKPRQNGQFVGVKGYM 544
Db 574 KDTAFTAGRALIYLPDNGFAFVSVYSTSPFSGPDGQNNPFKPTTGEQWEGVLKYPV 633
Qy 545 DDLNARVSVFPMKDKNAAPLNPNKKTRY--AALGKRVMEGVETEISGAVTPKQWIHA 602
Db 634 GYDALITVSADFLOKQNV-----PTYDEPTLPAQTGTEIHVQIGIEGKATVDFSLDLIA 688
Qy 603 GYSYLHSQIKTASNRDDGIFLLMPKHSANLWTTTQV---TPELTIGGV-NAMSGITS 657
Db 689 AASYTDSVYSKADDTGQGNKVRFPVNSLWKGKRIEDPGLGFGAGVGVHSSGGVD 748
Qy 658 SA-GMHAGGYATFDAMAAYRF---TPK---LKLQINADNIFNRHYARVGGANTFN--IP 708
Db 749 AANSFKYPATVVDAAISYDFGKQNPKEGLELNVTAQNLFPDKTY--VSGCSNINSCFY 805
Qy 709 GSERTWTANLRSF 722
Db 806 GKSRAVYANLSYKW 819

RESULT 31
AC2134
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2134
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
baena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2134
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-872 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074325.1; PID:g17131719; GSPDB:GN00179
C:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2626
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
ogy

Query Match 11.9%; Score 449.5; DB 2; Length 872;
Best Local Similarity 24.2%; Pred. No. 3e-23;
Matches 179; Conservative 120; Mismatches 303; Indels 139; Gaps 27;

Qy 47 IVTEKNGDYSFPAVTGKIPASLREIPQSVSIITNOQVKDRNV----- 90
Db 206 LVTGEQNRVYRPNASTGTRDTLIRIPQIVVPEQVQIDQVRTRLRDALLNTGGVVQD 265
Qy 91 ----DTFDQLARKTPGLRVLSSDDG-RSSVYVARGYSEYNIIDGLPAQMOSINGTLPNLF 145
Db 266 GGFSTSDQI-----GIRGFGDGTFGGSLVDGFK-----DGRGIRETAN----- 307
Qy 146 AFDREVNRGPGSLFDSSGEMGIVNLVRKRPYKAFQGHAAAGFGTHQYKAEADVSGSL 205
Db 308 -VERIEVLKGPASVLYGVQVPGGIVNLVTKQPLRDPYNAELSVGSFSTPRPSIDISGPL 366
Qy 206 NSDGSVGRVMAQTGASPRPAEKNNRHETFYAAAD-----WDINPDT-- 248
Db 367 NSDKTLLYRL-----NSVYETSDGFRDNQVQVREFISPTLKWEIGKATNL 412
Qy 249 VLGAGLYLQQRHLAPYN-GLPADANNKLPSPHQHVFVGADWNKFKMNSHVDVADLKHVFG 307
Db 413 TLQPDYLANDER---PDRGFLAFGEIIDT-PLERFFGEPDDVRKRVBEIGLSYLEHNFN 468
Qy 308 NGGYKVGMYSDRDADSNYAFAGSKLGKMKTPAGPGCNTADDKACAVGLGTEIKOKALA 367
Db 469 DNWKIRNAFRYQSSDT-PDYRAEPVRLNETTGILTRNFRSNDVYEENVTLTQDVVGK--- 524
Qy 368 FDASYRPPRLGNTANEFVIGADYNRFRSTNEQRTTLYARGGL---ALNEFRSI----- 419
Db 525 -----FITGSINHLLFLFGVDLAREFTSGTQSRLP-----GVITPSINVFNVPYNAIP 571
Qy 420 -POVDLIANARKGVRGYSHTVATENLDE---FGIYGKSTFHPADGLSLIGGRLGHYKIE 475
Db 572 RPDDELITN-----VVRNQDRSTGLGIFLQNTAFADNLKLLVGGRTFDVQDN 620
Qy 476 SGEKTLHKASK--TKFTGYAGAVYDLDNNSLYLSLSQLYTPQTNLADGKLLKPRQGN 533
Db 621 STDLRDGSSEGRYDSAFTRPLGIYVQPIEPIISLYASYSQSPQNFGRVDSGSIIEAERG 680
Qy 534 QFEVGYKGYMDDRLNARVSVFPMKDKNAAP--LNPNNKTRYAALGKRVMEGVETEIS 591
Db 681 QYEVGVKGEFLDGLDLAATLAAYHTIKNAIATDLNDPD---FLPLIGEQRNOGIELNVA 736
Qy 592 GAVTPKQIILHAGYLSHLSQIKTASNRDDGIFLLMPKHSANLWTTTQV---TPELTIGG 647
Db 737 GEISFGNNVTASYSHIDAETRDNDGLQGNRPANVPNTASFWTTYELOQGDLOGLFGFL 796
Qy 648 G---VNAMSGITSSAGMHAGGYATFDAMAAYRFTPKLQINADNIFNRHYARVGGANT 704
Db 797 GLFVVGDRQGDSSNTYI-IPGLRTDA-AIYKRDNRNAGINIQNLFNEXHYL---GANP 851
Qy 705 FNI---PGSERTWTANLRSF 722
Db 852 GRVAIEFGAPLTVIGSPSVTF 872

RESULT 32
AH2626
ferrichrome iron receptor Atu0409 [imported] - Agrobacterium tumefaciens (strain C58, Du)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AH2626
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH2626
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41430.1; PID:g17738752; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)

C:Genetics:
A:Gene: Act0409
A:Map position: circular chromosome

Query Match 11.7%; Score 441.5; DB 2; Length 708;
Best Local Similarity 25.3%; Pred. No. 8e-23;
Matches 191; Conservative 99; Mismatches 344; Indels 121; Gaps 32;

```

Qy 27 AACGADLETHVIGKRSYNALVTEKNGDYSFVAVTGTRKIPASLRIPQSVSIITNQV 86
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 8 AGCTVLKQITVEGQGENA--TGPRGVYAKKSATGSKTETETKALPQSVSVGQEMD 65
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 87 DRN-VTFPQIAKRTPLRLV--LSNDGSSVYARQYSEYNI--DGLPAQMSINGTL 141
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 66 DRCAVTKIEVELEKTYTTPGVAEPFGTDPDTDFWFIYRGFAQTQVTFDGLNLFSYGEFG 125
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 142 PNLFAFDVYVWRGPGSLPDSGEMGIVLVYKR-----PTKAFQ-----GHAAGF-- 189
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 126 MDAYGLERVEVLKGPASVLYGANGPGGIYGVMSKRAODTPVRETEIGINNFGNAFPGFDL 185
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 190 -----GTHKQYKADYSGSLN-SDGS--VRGRVAAQTVCASPPRAEKNRHETFYAA 239
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 186 GDKVDAEGVWK-YRVTGKYSGGNITDYSIDLAFIMPQ-ITFEP-DAQTSATLYGFFSA 242
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 240 ADMNDIPDTVLGAGTYIQQRHLPYNGLPADANNKLPSPFQHVFG-ADWNKFRMNSHDV 298
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 243 LD-----QVHVGNGFL-----PYGTAVDA--PFGKLDKKAFFYGEPPIDNGRVYQSMV 288
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 299 PADLKHYFGNGGKGVKGMRYSD-----RDADSNVAFASGKLGKMTPARPGC 345
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 GYDVSHFEDGKWKISQNAKRGHLYKHETGPYPCGMANADAN-----GQPIIL 334
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 346 N--TADDKACAVGLGTEIKOKALAFDASYSRPERLNTANEVIGADYNRFRSTNEGRT 403
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 335 DPTINDYMLTRFGYDGLSKVDSFGVNRLEGQDTGAVNHSPLFGIDYKRYRLDQVQACC 394
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 404 TLYARCGALNERRSIPQVDLIANARKGVG-----YSHTVATENLDEFGIYKSTP 455
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 395 GSNALGAL-----KPYVGYSTGTFVYVADNIVIQ--QOIGIYADQL 434
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 456 HPADGLSLGGGGLGHYKIESGE--GKTLHRSKTKFTGYAGAVVDLNDNNSLYLSLQ 512
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 435 RRGDGLVTLNGYVDVDTLNNRLPAGVSRSDALSGRAGLAEPFDGLTPYVSAT 494
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 513 LYPQTNLADGKLLKPRQNGQFVQ--YKGSYMDRLNARVSFYEMKDKNAAPLNPNN 570
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 495 FNNPLDITLADGTPAPSEHGFQEAIGKTPESFDSITASV-FKLVKD-NALVSTTAGG 552
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 571 KTRVYALGKRWEGVETISGAVTRKMOIHAGYSYLHSQIKTASNSRQDGF-LMPKH 629
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 553 VTTT--GQFQVESTGFELEKAKNLDENWKALASYSYTDLDITKDAHPNLIIGKSPWIPAH 611
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 630 SANLWTTQVTPR---LITGGCV--NANGSITSSAGMHAAGVATPDMAAAYFTTKKL 683
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 612 TSLWMDYVFTBETFEGLSIGGVRYQKSMADANTLRVSDAAYVDALRYE-KNDWTA 670
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 684 QINADNIFNRHYARVAGANTFNI PGSERTWTANL 718
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 671 SVNVANVFDEKEYKSCAGVSVCGM-GDSRTIITKL 704
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 33

Db3081
Probable outer membrane receptor for iron transport PA4514 [imported] - Pseudomonas aer
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text change 14-Sep-2001
C:Accession: D83081
R:Stover, C.K.; Pham, X.Q.; Brwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Laidig, K.; Lam,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950, M01D:20437337, PMID:10984043

A:Accession: D83081
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-753 <STO>
A:Cross-references: GB:AE004865; GB:AE004091; NID:g9950752; PIDN:AA607902.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4514
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 11.7%; Score 441.5; DB 2; Length 753;
Best Local Similarity 22.9%; Pred. No. 8e-23;
Matches 186; Conservative 129; Mismatches 318; Indels 179; Gaps 37;

```

Qy 15 ATVALASSVFPAQADLETHVIGKRSYN-----ATYENKGRYSFVAV--TVG 63
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 17 ASAGVATIRAIAPAAQDEA---GQKTDKDRVLSLDAATIVGQDETTYNVRSAS 73
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 64 TKIPASLRIPQSVSIITNQVKNRVDTFDLARTPLRLVLSNDGRSS--VYARGV 120
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 74 KKTATLDPKTVTVIPQVAKDGTALFLDALRTTPGITFGAGCGNPADRPPIRGF 133
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 121 E-YSEVNIQGLPAQMSINGTLNPLFAFPRVEMRGSGLPDSGEMGIVLVYKRKPTK 179
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 134 NMSDTFLDG---MRDVASQTRVEPVNQIEVSKGPGSAYTGAGSTGSLNLSK--TA 187
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 180 AFGCHAAAGF--GTHKQYKADYSGSLNSDGSVRGRVAAQTVCASPPRAEKNRHETFY 237
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 188 KQDNFTDAGFTMGSDQTRRTTLTDVNRMTGDNAFLNLN-----KIDAHV 232
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 238 AAAD-----WDINPDTVLG-----AGLYQQRHLPYNGLPADANNKL--PSLPQ 280
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 233 AGRDEVSVSRMGVAFVTGCTPRTATLSYHLSTDDMDPGLPLTNVNRSKANPSKA 292
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 281 HV-----FVG--ADNKKFKMS-----HD-----VFADLKHYGN--GGYG 312
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 293 SVDRNPFYGLKORDYKSTTDSGTFRIEHDLNDNLTLNSLTRVTRTLTDYIVSNPDSRG 352
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 313 KYGMRYSDDDADSNVAFASGKLGKMTPARPGCNTADDKA-----CAVGLGTEIK 362
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 353 NVANGVYVSAS-----RNSTSKGVNQDULKANETGCTIKTTLVTGL----- 396
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 363 QKALAFDASYSRPFRL-----GNTAN--EFVIGADYNRFRSTNEGRTLYARGLALNE 415
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 397 --EFSYEDVHNRPYAITSGGAGNTCNARLLASGQDTSLNRPFGDNTGSIITDGLATYD 454
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 416 FNSIPQVDLIANARK-----GVRYSHTVATENLDEFGIYKSTHPPADGLSLIG 466
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 455 TDTKTSAAVVPFTKLKSEQWEINLGLR-----YDDPDT--KSSGYQTAGR-- 498
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 467 GRLGHYKIESGGKTLHRSKTKFTGY--AGAVYDLNDNNSLYLSLQLYTP-----Q 517
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 499 GPAGYFKRE-----NSHFVNYQTGLYKRAPNSITLAWSTSNPFGETGEGEQ 548
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 518 TNLADGKLLKPRQNGQFVQYKGSYMDRLNARVSFYEMKDKNAAPLNPNNKTRVAA 577
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 549 AIDSVGNNDLDERNRNLELGTKMAFFDDALSLNALPR-TKTNARVASP-DVSLQVYL 606
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 578 LGRVMEGVETISGAVTRKMOIHAGYSYLHSQIKTASNSRQDGFILMPKHSAN--LW 634
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 607 DGEORVQGVLEFGNGLTEKMKVFGGYTLDEIRKSTVKSDEG--NKMPQTAQNNTILM 664
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 635 TTYQVTPBELTIGGG--VNANGSITSSAGMHAAGVATPDMAAAYRTPKLQIINADNF 691
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 665 TTYDLDQNTFTIGGTYVYKQYGNFANS--TYLPSTWRYPAMASYSKVSKNVDQLQNLVLT 723
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 692 NRHYARVAGANTFNI-PGSERTWTANLRYSF 722
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 724 DKRYFPQVYSTMAHVAAPG--RTALLGVNHFH 753
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 34

A63481

```

Db      659 VTSVDNGKLDAYVHHVNMVLGYQATRDILKQLQNLNLFDPKAYVERV 7033
      :||      |      |      |      |      |      |      |      |      |
RESULT 35
AE2085
ferrichrome-iron receptor [imported] - Nostoc sp. (strain P
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_ch
C;Accession: AE2085
R;kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamot
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yana
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitro
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE2085
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-858 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA073935.1; PID:g1713
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: al12316
C;Superfamily: ferrichrome-iron receptor 1; tonB-dependent
      Query Match      11.2%; Score 424; DB 2; Length
      Best Local Similarity 23.8%; Pred. No. 1.8e-21;
      Matches 170; Conservative 124; Mismatches 328; Inde
Qy      47 IVTEKNDGYSFSAVTGKTPASLREIPQSVSIITNQOVKDRNVDTFDQ
Db      199 VVTGEQDTYRVREATTATKLDVPLRDVPASVQVIPKEVIQDRQVIRLNE
Qy      107 SNDDGRSS--VYARGYYS--EYNDIGL-----PAQMSINGTLPLNLF
Db      259 ETYGGLASQGYFIRGFGTGPETLDRGDFRDFGIFSPRDVANT-----
Qy      157 -SGLFDSSGEMGGTVNLVRKRPTKAFQGHAAAGFTGHQYKAEADVSGS
Db      309 ASVLXGSAFSGGVVNTITKKPLADPFYQINGTIGNDFYKSSIDFSGP
Qy      216 MAQTVGA-SPRPAEKNNRHETFYA-AADWDINPDTVLGAGLYQOORHLA
Db      369 NAAVENARSPDFVEN--ESTFIPIITVKAGERTNLTFGYEYKYDYT
Qy      274 KLP-SLQPHFVG-ADWNKFKMNSHDVADLKHFGNGGYGKGVGMRYSD
Db      425 KVPFDLPINRFLGEPNLNRGEFKSNFTYLTLESEFGYNNNNKKPQAFNV
Qy      332 SKLGMKTPAGRPGCNTADDKACAVGL-GTEIKOKALAFDASYSRPFRLG
Db      480 -----TKGAQPDINADGQTVARGYNVDRAQNNLSFONEISKFNFTG
Qy      391 YNRPRSTNEQRTLYARGGLA----LNEFRSIPQVDLIANARKGVRCY
Db      534 LSYXKYSVD-----FLSGISGLDIDFNPVVGAPAPDLSRSNEEYGG-
Qy      447 FGIYKSTFTHPADGLSLIGGRIG-----HYKISGEKTLHKASKTKPF
Db      578 IAVYFQNLVELTPQIKLLAGRGFTVDSVVRDVT--GTVVYDETSDFSXF
Qy      502 DNNLSYLSSLQLYTPQI-NLDAADGKLKPRGNGFVGYKGSYMDRLNL
Db      636 DSTSIYASWNSFNPQIFGKTRNNEPKPETAQEFVGIKQEFKLNRKUS
Qy      561 NAAAPLNNPKKTRYA-ALGKRVMGEVETEISGAVTPKWQTHAGYSYLH
Db      696 NV---LTTDPEDDNFNSIQTEQKSRGVELDVVGELPCWKLIIGTYTYD
Qy      618 ---RDDGFIILMPKHSANLWTTYQV----TPELITGGVNMAMSGITSSA
Db      753 VNNRLVG----PYNASLWTTYBELQGNLQGLGLGVVYVVGREASL

```

QY 669 FDMAAAYRTPKLTQINADNIFNRHYARVGANTNIGSERTWTANLRYSF 722
 Db 809 TDSIFPK-RDNRAALNFKLPTKTYE---SQSFIVPAAPFTVIGVSFEF 858

RESULT 36

AC0547
 ferricxamine B receptor precursor [imported] - *Salmonella enterica* subsp. *enterica* serov. *C*/species: *Salmonella enterica* subsp. *enterica* serovar *typhi*
 A/Note: this species has also been called *Salmonella typhi*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C/Accession: AC0547
 R/ParKili, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Comerford, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moutle, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov. *C*/Accession: AC0547
 A/Reference number: AB0502; PMID:11677608
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-696 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD08819.1; PID:g16501633; GSPDB:GN00176
 C/Genetics: A/Gene: STY0396

Query Match 11.2%; Score 421.5; DB 2; Length 696;
 Best Local Similarity 23.1%; Pred. No. 2e-21;
 Matches 175; Conservative 121; Mismatches 353; Indels 109; Gaps 25;

QY 12 MTAATVLAAL-----SSVFPAQTADLETYHIKQGRSNAIVTEKNGDYSSFAVTVG 63
 Db 1 MFATRRALLIGALIGATFPLFAQETTKNDTV-----IVT-----SPVSGA 43
 QY 64 TKPASPAREIPQSVSITTNQVKDRNVDTQDARKTPGLRVLSNDGRSS---VYARG 119
 Db 44 TRLATPDIETPOSVSIITRQOPEOGATSVRAQVSYTPG---VYSNQIGANRFDYILRG 101
 QY 120 YERSEVN---IDGLP---AQWQINGTLPLNLPAPRVVEMVSGSLPSSSEMGIYVLR 174
 Db 102 FSGSGLDNVTLDGKMKMGJTNSSSLVADPWPLEDIEVGPASVLYGRSSPGIYSLTS 161
 QY 175 KRPTKAFQGHAAAGFTGHQYKAEADVSGLNSDGSYRGVMAQTVASPRPAEKNNRHE 234
 Db 162 RKAFADAGSEVKLPAQNNNGRGAAPDVTGPDLDNERAALSGMTRKYADQCFPLKEERY 221
 QY 235 TTPAADWDINPVTVG-AGLYVQ-----QRHLAPYNGLPADANNKLPSPLOH 281
 Db 222 ALMPSLTWRITDRTRLDLMAVYLHRDPQGGSHGLPYQGTAVVPYNG-----GKISNT 272
 QY 282 VPFGA-DMNKFKNNSHDVPAADLKHYPENGSGYGVKMYSPRDADSNAPAGSKLGMKTPA 340
 Db 273 FFGEDDYDYDRRENWVGNIEHLFENGWSVQKLELYLTKYTLQVYAAGWLN----- 327
 QY 341 GRGCGTADDKACAVGLGTEIKOKALAFDASYSRPFRLGTANEFVGAAYNFRSTNEQ 400
 Db 328 -----ETALNRGYS---GSEKMSATLADNQLDQSVDTGLNRLHVLGDIYQ-----DRS 374
 QY 401 GRTTLVARGGLALNFRSIPQVDLIANARKVGRGSHVTATE--NLDFEIGYKSTFHPA 458
 Db 375 NHTTGV-----YGAFPPIDAFNPVGAQPDYITLYREKHKLRQGTUYIQQDM-SW 424
 QY 459 DGLSLGGGRLGHYKTESGGKTLHKA---SKTKFTYAGAVYDLNDNNSLYLSLSQLY 514
 Db 425 DRRRFTLGGYDRVSNSND--KLHDSRSLDKNNVSTRALLYLFPNGVAPYLSYSTAR 482
 QY 515 TPQTNLDADKLKPRQNGFEVGYKGYMDRLNARVSFYRKKDKRAAALPLPNNKTR 574
 Db 483 TTSFADENQNVLEPMKQKQWEGAKTEPIGSGNSQPSAAYRINQNTIATKEEPTD--- 539
 QY 575 YALAGKRWEGVETEISGAVTPKQIHAGYSYLHSQIKTASNSRDGIFLLMRKHSANLW 634

Db 540 YRSIGETIESKGVLEIAISHLSDSVRLQAAATYTDIRYKSSPOEGKRAVAYAPRQAASAW 539
 QY 635 TTQVQTPRE-----LTIGGVNVMAGITSS-AGMEA-GGATFEDMAAYRFP-----KXKQ 684
 Db 600 LSTDVKSGLEMLITLDSGIRYVGVTSDRLNTHTLSYTLVDVWVGVDLSIGLNSAQ 659
 QY 685 INADNIFNRHYARVGANTNIGSERTWTANLRYSF 722
 Db 660 LNVNMLLDKRYVAACNSL-SYCFGAERSIVGSVNAF 696

RESULT 37

AC0209
 ferrichrome-iron receptor [imported] - *Nostoc* sp. (strain PCC 7120)
 C/Species: *Nostoc* sp.
 A/Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C/Accession: AC2079
 R/Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*
 A/Reference number: AB1807; M0ID:21595285; PMID:11759840
 A/Accession: AC2079
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-853 <KUR>
 A/Cross-references: GB:BA000019; PIDN:BAB73884.1; PID:g17131276; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics: A/Gene: alr2185
 C/Superfamily: ferrichrome-iron receptor 1; lonB-dependent receptor amino-terminal homol

Query Match 11.1%; Score 420.5; DB 2; Length 853;
 Best Local Similarity 23.1%; Pred. No. 3.1e-21;
 Matches 187; Conservative 117; Mismatches 327; Indels 177; Gaps 30;

QY 3 QPMSVRINMTATVL--AALSSS---VF--AAQTADL-----ETVHIKQGRSYN 45
 Db 135 QDANTTRVSIIGETITLPTHELFPDSDEGLIFGLTSVTSASLPQOPEQSESPAAEPIE 194
 QY 46 AIYT-EKNGDYSSFAVTGKTPASLREIPQSVSITTNQVKDRNVDTQDARKTPGLR 104
 Db 195 LVVTGKDGQVQNTATV--TRTWTPIIDIPQIQVPRQVLEQQTTRVDALANRPGVL 253
 QY 105 VLSNDGRSSVARGYEVSEYNI--DGLPAQWQINGTLPLNLPAPRVVEMVSGSLPDS 162
 Db 254 GSTNAETIGNQITIRGFSTSNLPILRDGFRIYE---NFSQETSNLRIEVLKGPASVQYG 310
 QY 163 SGEMGIYVLRRLPKAFQGHAAAGFTGHQYKAEADVSGLNSDGSYRGVMAQTVGA 222
 Db 311 QLDPGVINYLVTKPLSEPPYEIOAQFGSGYGLRPSFDVSGPLTDGKLYRLNA--TYOR 369
 QY 223 SPRAEKNRKHETFYAAD--WDINPDTVL--GAGLYVQQRHLAPYNGLPADANNKLPSTL 278
 Db 370 EBFRRFPNTEREFPLAPSLTWMKISDRTNVDSLETLDSTR---PDTSLVAFGRVADV 426
 QY 279 P-OHVFVADWNNKFKNNSHDVPAADLKHYPENGSGYGVKMYSPRDADSNAPAGS----- 332
 Db 427 PYSRVENDPD-DEIDKFSIAVNLHRRFSDWTLTNSFRYLOQDLFTQATLAGSINETT 485
 QY 333 KLGMRTPAGRPGCNTADDAACAVGLGTEIKOKALAFDASYSRPFRLGTANEFVGAAYN 392
 Db 486 GIITRTYAOH-----EYKSDYSLQTNVVGKFTTGSIKHTLLTLAGVDN 528
 QY 393 R-----FRST-----NEGRFTLYARGGLALNEF 416
 Db 529 RGLLDLVLRGRFTTININPVYGVPRPTDFTLPRATPFKKTETRLRGYLDQDIALNN-- 587
 QY 417 RSTIPQVDLIANARKVGRGSHVTAVENLDFEIGYKSTFHPADGSLIGGGRLGHYKIES 476
 Db 588 ---QFTVLAGLRVDTVDKDFETDESKYD-----SAMSPLYGL----- 622

```
Qy 477 GEGKTLHKASKTKFTGYAGAVVLDNNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQPE 536
Db 623 -----VYKPVENLSIYTSRSFVPSFRDANGDFLQPERGAGYE 662
Qy 537 VGYKGSYMDRLNARVSFYRMKDQNAAPLNPNKKTRYAALGKRVMEGVETEISGAVTP 596
Db 663 VGKIAELLOGNLFATLAYFDITKQNVST-ADPDVLGAS-VATGEQRSGRIELSVAVGTIAP 720
Qy 597 KWQIHAGSYLSHQIKTASNSDDGIFL-----LMPKHSANLWTTYYQV-----TPELTIGG 647
Db 721 GWNIIAGYAYTDAEI-----SKONTIAGVNRLLPGTPKHSANLWTTIYEQKSLQGLGFL 775
Qy 648 GVNAMSGITSSAGMHAGGVAATDAMAAFT-----PKLKLQINADNIIENRHHYARV 699
Db 776 GNVYV-----GKRFNGQNDFEVDSYFLTNAALFYRQNNRVGLNFNNIFDINY---I 825
Qy 700 GGANTFN-----IPGSERTWTANLRYSF 722
Db 826 SSAALTRTRSTEPGQPFVTVGSISLEF 853

RESULT 38
S74457
ferrichrome-iron receptor 3 - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1490
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Accession: S74457 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74457
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <KAN>
A:Cross-references: EMBL:D90899; GB:AB001339; NID:G1651650; PTDN:BAAL16609.1; PID:G165168
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: fhvA_3
A:Start codon: GTG
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
F:241-371/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:579-853/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 11.0%; Score 416.5; DB 2; Length 853;
Best Local Similarity 23.2%; Pred. No. 5.9e-21;
Matches 178; Conservative 116; Mismatches 311; Indels 161; Gaps 27;

Qy 10 INMTATVLAALSSVFAQTADLETVHKQGRSNAIVT-RKNGDYSSFAVTGTPKA 68
Db 165 VNATQSLVLGLSPGKVADEED-----GNDAIQVVVVGSDREGYAVDDATTATLTD 217
Qy 69 SUREIPQSVSIITNOQKDRNVDTFDQLARKTPGLRVLSNDDGRSSVYARGVEYSEYD 128
Db 218 PURDIPQSIQVVPQVLEDRQIRASEALQNVSGVQRGNVTGCTSEI-----FNIR 268
Qy 129 GLPAQMQSINGTLPNLFAF-----DRVEYMRGSPGLFDSGEMGGIVNLVRKR 176
Db 269 G----FQPGGTLRDGFKFRDNFSIPTANTLQRIEVLKGPASVLYGNLDPGGVINVTQ 324
Qy 177 PYKAFQGHAAAGFTHKQYKAADVSGSLNSDGSVGRV-MAQTQVAGSPRAKNNRHET 235
Db 325 PLSEPFYEVAMQAGNGLVRPTIDLSGLPLNSQRTALYRLNAAVEGGNFRDPTDEVARFF 384
Qy 236 FYAAADWDINPTVLGA--GYLYQORHLAPYN-GLPADANNKLPSLPQHVFVGADWNKFK 292
Db 385 ISPVVTWQLSDQTLRFEWDYLYDR--PFRGIVA----- 418
```

```
Qy 293 MNSHDVFPADLKHFGNG-----GYCKVGMRYSDRDADSNYAPAGSKLGMKTPAGRPGCNTA 348
Db 419 -----EGTGADIPFDFRVLGELDFDARTNFS-AGYRLRLEHR-----FS 455
Qy 349 DDKACAVGLGTEIKQKALAFDASYSRPRL-----GNTANER-----VIGA 389
Db 456 DNWK-----LNRFRPSYLDQAAEQTELVRDDETTGNLSRQFSRNEQQIRNVELQTDLLIG- 510
Qy 390 DYNFRSTNEOGRITLYA-----RGLLALNEFRSIPQVDLI-----ANARKGV 432
Db 511 ---KF-STGPIQHTLLFGVDLSWQSAPPIPRGVA-----APTINIFNPVYGVITAPSI 560
Qy 433 RYSHTVATE-NLDFEIGYKSTFHPADGLSLGGRLGHYKIESGEGKTLHKASKTKFT 491
Db 561 NDFPDEVSSSEQTTGLFDQDQVTLTDNLKLMGGRFDTIDQSSSSNGESDERYDQAFS 620
Qy 492 GYAGAVYDLNNDNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQFEVGYKGSYMDRLNAR 551
Db 621 PRLGLIVQPIEPVSLYASFSSRFQNFGRFDGSLLEPVFTQYEVGVRGEFLDGRLIAN 680
Qy 552 VSFYRMKDKNAAA--PLNPNKKTRYAALGKRVMEGVETEISGAVTPKWQIHAGSYLSHS 609
Db 681 LAAYEITVSNLAVTDPENPFS-----IPSGEQRSGVEFDIAGEILPCWNIIASVATDA 736
Qy 610 QIKTASNSRDDGIFLLMPKHSANLWTTYYQVTPELTIGGVNAMSGITSSAGMHAGGVATF 669
Db 737 RVTKDDNLEPKNLLEGGVFPFNSASLWSTYEQ-----AGDLQGLGFLGLFVYGERQG 788
Qy 670 DAMAAYRFTPKLKLQINADNIFNR--HYARVGGGANTFNIPGSERT 713
Db 789 DLNNSFQIPSLYRTDI---SVFYRRNNWRAAINVNLNFDIDYEAT 831

RESULT 39
AC2132
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2132
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.;
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2132
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-857 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW74309.1; PID:G17131703; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2610
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 11.0%; Score 415.5; DB 2; Length 857;
Best Local Similarity 23.6%; Pred. No. 7e-21;
Matches 174; Conservative 106; Mismatches 296; Indels 161; Gaps 28;

Qy 47 IVTERKNGDYSSFAVTGTPKIPASIREIPQSVSIITNOQKDRNVDTFDQLARKTPGLRVL 106
Db 204 VVTQGDQRYSPNATATTATDTPLRDIPQSIQVAPRVQLEDQVIRASDALRSVSGVQQG 263
Qy 107 SNDDGRSSVYARGVEYSEYDNLGLPAQMQSINGTLPLNLF-----AFDRVEYMR 154
Db 264 NKVCGTSEV-----FNIRGFP-----QFGNLRGDFNNRNNFSIVETANLERIEVLK 310
Qy 155 GPSGLFDSSGEMGGIVNLVRKRPTKAFQGHAAAGF--GTHKQYKAEADVSGSLNSDGSVR 212
Db 311 GPASVLYGNLDPGGVINFTVKQPLS--EPFYAAGLQVGSFGLVRPTLDSLGPLNPERTLL 368
Qy 213 GRVMAQTQVAGSPRAEKN---RHETFYAA--ADWDINPTVLGAGLYQORHLAPYN-G 266
```

```

Db 369 YRLNA-----AYEBCGNFRNFDTEVERFFISPVVTWKIGDRTDLRLLEY-SNDKRPYDRG 423
Qy 267 LPADANNKLPSPQHVFAVDWNNKFKANSHDVEAD-----LKHFGNGGYGKVKR 317
Db 424 LVAFGNG-IADIPFDVLG-----BEDDSEKRTNPLAGRLHRFRPDMKLANQR 473
Qy 318 YSDRDADSYAFAFGSKLGMKTPAGRPCNTADDKACAVGLGTEIKOKALAFDASYSRPR 377
Db 474 YSSSDQTN-----RLEBGRINLETG-----LRRFES 501
Qy 378 LGN-TNEEVIGADYNRFRSTNEQRTTLA-----RGGLALNEFRSIPVDLIANAR 429
Db 502 LBEFTYRNRELQTDLVGNFATGSIQHTLLFGVDLSWVNDGVSLL-FEAPASIN-IFNPV 558
Qy 430 KGVRYGSHVATEENLDEF-----GIYKSTFHPADGLSLIGGRGKHIES 476
Db 559 YGL-----ATRPBDEFPADVFPFGSQDTSIGVFONQITLAEMLKLLVGRFDNDIDSS 612
Qy 477 GEGKTLHKASKTKFTGYAGAVYDLDNNSLSYLSQLYTPQTNLDADGKLLKPRQNGPE 536
Db 613 ASDERDQA-----FSPRVGVIVQPIEPISLYTSFSRSGFNFGNADGSLLEFVRGQYE 668
Qy 537 VGVKGSYMDRLNARVSFRMKDKMAA--PLNPNKKTIRYALGKRVWEGVTEISGAV 594
Db 669 VGVRGFEFLNGSLITNLAAVEITRSNLAVTDPDNPNS---IPSGQRSGRGLDVTGQI 724
Qy 595 TPWKQIHAGYSYLHSQIKTASNSRDGIFILMPKHSANLMTYQVPELITGGCVNAMS 654
Db 725 LFGMNLIAAYATIDAVTDADDNLQPGNLADGVFPNSASLSTYEIO-----TGDPQG 776
Qy 655 ITSSAGMHAGG-----YATPDMAAAYFTPEKLQINADNIPNRHYA-- 697
Db 777 LFGGLGLFYVGEHQGLNNSFQLPSYVRTDASIFYR-RNNWRAGININMLFNVDIEDSG 835
Qy 698 ----RVGANTPIRPGS 710
Db 836 QRRNRINPGEFPTVRGT 852

RESULT 40
C95382
probable ferrichrome-iron receptor [imported] - Sinorhizobium meliloti (strain 1021) mag
C|Species: Sinorhizobium meliloti
C|Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C|Accession: C95382
R|Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barclay-Hubler, F.; Bows
., Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A|Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A|Reference number: A95262; MUID:21396509; PMID:11481432
A|Accession: C95382
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-714 <KUR>
A|Cross-references: GB:AE006469; PIDN:AAK65621.1; PID:g14524105; GSPDB:GN00165
A|Experimental source: strain 1021, megaplasmid pSymA
R|Galibert, F.; Flman, T.M.; Long, S.R.; Phihler, A.; Abola, P.; Ampe, F.; Barclay-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federpiet, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A|Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Kles, E.; Komp, C.; Lelaure,
hebaull, P.; Vanderbol, M.; Vorholter, F.U.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A|Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A|Reference number: A96039; MUID:21368234; PMID:11474104
A|Contents: annotation
C|Genetics:
A|Gene: Sma1747
A|Genome: plasmid

Query Match 10.9%; Score 411; DB 2; Length 714;
Best Local Similarity 24.6%; Pred. No. 1.1e-20;
Matches 187; Conservative 110; Mismatches 344; Indels 120; Gaps 30;

```

```

Qy 12 MTAATVLAALSSSVFAAQ-----TADLETYHIKQORSYNAIVTEKNGDYSSFAVTGT 64
Db 24 LIGCTAFALTLALSPAQDAVPEGDTTVLETIVAHGAGGSVLTNDEDSKSIATRTGA 83
Qy 65 -KIPASLEIPQSVSITINQOYKDRVDFDQARTPL--RVLNSDGRSSVARGVE 121
Db 84 GKMPFDILAPASVSVITSKIEERAAADTIEQVQVTAQVVDPFYSDRYDYDFDRGPT 143
Qy 122 YSEFINIDGLPAQMOSINGTLPLNLPADRYVWRKPGSLFDSGEMGIYNLVKRPKTKAF 181
Db 144 PTYTR-DGL-AIGRTPAVGRPEEYAFERILEVKGASSSFGAEPGGSVNYTKTKPSR 201
Qy 182 QGHAAGFPTHKOYKAEADVSGSLNSDGSVRGVAQTGASPRPAEKNNRHETFYAAD 241
Db 202 FGEVYGTGGSFSHKELGDFGDNLTADETL-----SYRLGKQORSQAEVDYDQ 250
Qy 242 WDN-----PDTVLGAGTLYQOQRLAPYNGLPADANNKL-PSLPQHVFGAD--W 288
Db 251 DDENFVWGVTWRPTDASTLTFIFD--HL-DKDGVPGSGHPLGTDPFDRDQFGEEDYYF 307
Qy 289 NKRKONSHDVPADLKHYPGNGGKVKWYSDRDADSNYAFAGSKLGMKTPAGRPCNTA 348
Db 308 SETNRSISVLPD--HDFGNGLSFPSSNARYSNLNDGFGSAVIGS-----TP 351
Qy 349 DDKACAVG--LGTETIKOKALAFDASYSRPFPLGNTANEFVIGADYNRFRSTNEQRTTL 405
Db 352 TGGSYTAGRYFFPGENKSTQGVFIDALVYEAALDNVESTTLTGADYNKESD----- 403
Qy 406 YARGGLALNEFRSIPVDLIANARKGVRGYSHTVATENIDE--FGIYKSTFHPADGLSL 463
Db 404 -----SANFYAPAPSIDMEDPIYSGGPGAMA PYASTNNDOQTNAIYLQODLTFEFDKLT 457
Qy 464 IGGG-----LGHYKESGEKTHKASKTKFTGYAGAVYDLDNNSLSYLSQLYTPQT 518
Db 458 SPGLRNDWLDLSETNLAG--TRRAGNREFTYTRIGASYKTEELAPYISAESAP-- 512
Qy 519 NLDADGLKLPKQNGQFEGYKGSYMDRLNA--RVSFRMKDKN-----AAAPLNPNK 571
Db 513 --PAAGS--DPTTGQYEVGIR--YRPDAFPAMFTASVVDLTKGNTVDPQVYLYPQYVE 566
Qy 572 KTRVYALGKRVWEGVTEISGAVTPKQIHAGYSYLHSQIKTASNSRDGIFILMPKXSA 631
Db 567 KYRH-----RGFELKAKAEVTNNISVIAAYSIDSICEFGCANDGNRLMRVPKMA 618
Qy 632 NLMTYQVPELITG--GVA-----MSGTSSAGMHAGYATFDMMAAYRFPKXK 682
Db 619 SWGTYTLLEGDARGGMLFGIARVTDAYTISITNTSSEA--VVFDAAFYTKIQENTT 676
Qy 683 LOINADNIFN-RHYVAVGANTFNPISGSERTWTANLRXSF 722
Db 677 FQLNVNMLFDEKHVASKDSGAVYYN-PG--RSILATLRQSW 714

RESULT 41
C81861
hypothetical protein NMA1663 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C|Species: Neisseria meningitidis
C|Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C|Accession: C81861
R|Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jags, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A|Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A|Reference number: A81775; MUID:20222556; PMID:10761919
A|Accession: C81861
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-635 <PAR>
A|Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84891.1; PID:g738030
A|Experimental source: serogroup A, strain Z2491
C|Genetics:
A|Gene: NMA1663

```


Query Match 10.8%; Score 407.5; DB 2; Length 635;
Best Local Similarity 24.4%; Pred. No. 1.6e-20;
Matches 158; Conservative 111; Mismatches 256; Indels 121; Gaps 26;

QY 8 FRINMTAAATVLAALSS-----VFAAQTADLETVHIKQORSYNAIVTEKNGDYSSFAV 60
DB 22 FSLKILTVMLLSAYGSPADGVVPSDGNVTSLDVTNVRGSHALSG-KTEKTRSTYIDRM 80

QY 61 TVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQLARKTPGLRLVLSNDGGRSSVARG 120
DB 81 STATGMRIAGKTPQSVSVITSRLLDDKAVHTLEAMKMTTGVNVVRDSSGLQTRFLSRGF 140

QY 121 EYSEYNIDGLPQMOISING-----TLPNLPAFDRVVEVMRGPGLPDSSEMGGIIVNL 172
DB 141 YIDQIGEDGITVNVAGRSYATAKIDVSPSTD LAVYDHIIEVVRGATCLTQSNSEPGGTVNL 200

QY 173 VRKRPTKAFQGHAAAGFGTHK-QYKAEADVSGLSNDSGVRGVMAQTVGASP-RPAEKN 230
DB 201 IRKRPTASPK-HTGELTADHRGSRRAVLVDVSGLSNKNANTLRGL-----VGABEYKKSFKD 255

QY 231 ---NRHETFYAAADWDINDPTVLGAGLYQQ-----RHLAPYNGLPADA 271
DB 256 RVWGRKHMVYGTAEADAGDSSVLTGCMYKSRVDPDFSGIILSCENQKTAPFSSTP--A 313

QY 272 NNKLPSLPQHFVFGADWNKFMKNSHDVADLKHVFGNGYGVKVMRYSRRDADSNYAFAG 331
DB 314 CNRPLQPLRNTVLGDEWLSADKYNLFSGFKHVFDPNGQWLNQNAEVSYTKNESD----- 366

QY 332 SKLGMKTPAGRCPCNTADDKACAVGLGTEIKQKALAFDASYRPRFL-----GNTANEFVI 387
DB 367 AKVGOFPLKNEHAAGLSDED--AVGFLTE-KNEVIPFEPKQKALEKLKAYRDETAKEYNE 423

QY 388 GAD---YNNFRSTN-EQGRITLYARGGLALNEFRSIP-OVDLIANARKGVRSYHTVATE 442
DB 424 RKDDFVKNRFDNATFEQYSRRAERKAGFDCMSAPFALDFICQSWGDPGVD----- 477

QY 443 NLDEFIYKSTFHPADGSLUSLGGGRGLHYKESGKTLHKASKTKFTGYAGAVYDLND 502
DB 478 -----ADKSEFVD-----KALAKEGIFENNAQR-----FPNSLYDSGF 510

QY 503 NNSLYLSQLYTPQTN-----LDADGKLLKPRQNGQFVGYKGSYMDRLNAR- 551
DB 511 NRKATANRRYSYMPURHTKDDQWGIKLDLTGYGLFGREHDFFGY--AYGDEKIRSEY 568

QY 552 VSFYRMKDKNAAAAPLNPNKTRYYAALGRVMGVEVETI-----SGAV 594
DB 569 LEIYERRHR-----VRPNTGATHGVYAGS--CQGSRTVICLLMWSGI 609

RESULT 42
S74447
ferrichrome-iron receptor 1 - *Synechocystis* sp. (strain PCC 6803)
N:Alternate names: protein sll1409
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74447
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74447
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-863 <KAN>
A:Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIDN:BAAL16599.1; PID:g165167
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: fhuA_1
A:Start codon: GTG

C;Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
C;Keywords: iron transport
F:240-375/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:581-863/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 10.6%; Score 402; DB 2; Length 863;
Best Local Similarity 22.2%; Pred. No. 6.3e-20;
Matches 171; Conservative 127; Mismatches 335; Indels 136; Gaps 31;

QY 9 RINMTA-ATVLA-----LSSSVFAAQTADLETVHIKQORSYNAIVT-EKNGDYSSF----- 58
DB 150 RITVNGTGTVPAAQVIPSSENLILSLTPINTV--ESEEEIEIVATREEEAAVQEFFVPN 207

QY 59 -AVTVGTHKIPASLREIPQSVSIITNQVKDRNVDTFDQLARKTPGLRLVLSNDGGRSSVYA 117
DB 208 TSVATGTDTP--IMDTPFSAQWSEEVIRSQOAITLEDVLTVNVSSTFGGTTGGRETIFG 265

QY 118 -RGY--EYSEYNIDGLPQMOQSIN-----GTLPNLPAFDRVVEVMRGPGLPDSSEMGGI 169
DB 266 IRGFGNQFS-----DVTPLRDRGFRLYGGFQGITVSHLQOQVEVLKGPSSILYQIEPGV 321

QY 170 VNLVRKRPTKAFQGHAAAGFGTHKQYKAEADVSGLSNDSGVRGVVM-AQTVGASPRPAE 228
DB 322 INLSKKPLNEPFAEVEVOLGNOGLVRPRFDISGLAPSGNLYRLNCGVYNEASFRDPN 381

QY 229 KNNRHETFYAAADWDINDPTVLGAGLYQORHLAPYNGL-----PAD 270
DB 382 QPLERFAYAPIVTVAITDDTDLSLAVEYINDTNPADFLGSSFGDGVAPVPRSRVINDPSD 441

QY 271 ANNKLPSLPQHFVFGADWNKFMKNSHDVADLKHVFGNGYGVKVMRYSRRDADSNYAFA 330
DB 442 IVNK-----NFIISAGYN-----LEHFRFENWKLNRNFRYMSYNDYVNV--- 479

QY 331 GSKLGMKTPAGRCPCNT--ADDKACAVGLGTEIKQKALAFDASYRPRFLGNTANEFVIG 388
DB 480 ---IAPLPIVNGPTVTRFFADQG-----CQGSVSFYTNVAGKFESTGVKHELLAG 527

QY 389 ADYNRFRSTNEQGRITLYARGGLALN---EPRSIPQVDLIANARKGVRSYHTVATENL 444
DB 528 IDYN-----WSEESILTLFG-QPTSINVPDPDYNAIPKN-----RSDLPLFGDTFTSSN- 576

QY 445 DEFYIGKSTFHPADGSLUSLGGGRGLHYKESGKTK-LHKASKTKFTGYA-----GAVY 498
DB 577 -RLGYLDQVQVLSLENLILVAGLRDYDTITQNTNQLQDFNOGGNTQOQDSAVTRPRIGLLY 635

QY 499 DLNDNNLSYLSQLYTPQTNLDADGKLLKPRQNGQFVGYKGSYMDRLNARVSFYRMK 558
DB 636 RPIPEISFPFSNYSQSPFNSGIDISGNPLEPERGEGEIGYKAELEFQQLLTLLTYFNIS 695

QY 559 DNAAA--PLNPNKTRYYAALGRVMGVEVETISGAVTPKQIHHAGYSYUHSQI--KTA 614
DB 696 KNNVAVSDPVNPLFLST-----IGTQSQGIELDIVGEILPGWKIIGNYSYINAKVTEDTD 751

QY 615 SNRRDDGIFLLMPKHSANLWTTQVTPELTIGGGVNMASGITSSAGMH----- 662
DB 752 PNFVDNRFLGI-PYNMANLWTTYEIQS-----GALQGLGFGIGFNGYVDRFGDLANT 802

QY 663 --AGYATFDAMAAAYRFTPKLKLQINADNIFNRHYARVGGANTFNIPG 709
DB 803 YTVGDYIIGNAAFIYQ-RDKYRVALNLRNFNTNANYVRAVSGNQTGIBPG 850

RESULT 43
S22673
ferrinoxamine receptor precursor - *Versinia enterocolitica*
C:Species: *Versinia enterocolitica*
C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 29-Sep-1999
C:Accession: S22673
R:Baeumler, A.J.; Hantke, K.
Mol. Microbiol. 6, 1309-1321, 1992
A>Title: Ferrinoxamine uptake in *Versinia enterocolitica*: characterization of the receptor
A:Reference number: S22673; MUID:92349959; PMID:1640832
A:Accession: S22673

RESULT 45
E9118
probable ferrichrome iron receptor precursor [imported] - *Escherichia coli* (strain O157:H7)
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: E9118
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic
A:Reference number: A9629; MUID:21156231; PMID:11258796
A:Accession: E9118
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA837340.1; PID:gl3363390; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: ECs3917

Query Match 10.4%; Score 391; DB 2; Length 713;
Best Local Similarity 23.4%; Pred. No. 2.8e-19;
Matches 180; Conservative 127; Mismatches 325; Indels 138; Gaps 35;

Qy	20	ALSSSVA-----AQTADLETVAHKQRSYNAIVTEKNGDYSFPVTCVKTPASL	70
Db	15	ALFLLFAAPMIHATDTATTKOGSTITVTADAN---TATEATDGYQLSTSTAILTDMEM	71
Qy	71	REIPQSYSIITNQOQKORNVDTFQLARKTPGLRVLNSDDG---RSSVARGYSESYNID	128
Db	72	LDIPQVNTVSQVLEQNQATTLDEALYVNSV-VQINTLGGTQDAFVRGF---GANED	127
Qy	129	GLPAQMOISGTTLPNLF--APDRVEVMRGPBSGLFDSSGEMGGIIVLVRKRPYKAFQCHAA	186
Db	128	G-SIMTNGRLTVLPRSPNAATERVEVLKGPASTLYGILDPGGLINVVTKRPEKTFHGSV	186
Qy	187	A-----GFGTHQYKAEADSVGSLNSDGSVRGRVMAQTVGASPRPAEK---NNRHETFY	237
Db	187	ATSSSFGGT-----GQLDITGPTEGTQLAYRLTGEVQDEDEYRNFCKERSTFI	235
Qy	238	A-AADWDINPDTVLGAGLYQOR-HLAPYN-GLPADANNKLPSPLOHFVVGADMNKPKVN	294
Db	236	APSLTWFGDNATVT---MLYSHRDYKTFPDRGTIFDLTKQP-----VNVVD-RKIFED	284
Qy	295	SHDVFADLKHYPFGNGYKVGKVMRY---SDRDAOSNYAFAGSKLGMKTPAGRPCCNTADDK	351
Db	285	EPFNITD-----GOSDLAQLNAEYHILMSQWTAERDYSYODKY-----SDNQ	326
Qy	352	ACAVGLGTEIKQKALAPDASYPFRFLGNTANEFVGADYNNRFRSTNEQGRITLYARGGL	411
Db	327	ARVTAYDATTTGLTRRVYDATQGSFORMHSTRADLQGNVDIAGF-YNE-----ILGGV	377
Qy	412	ALNFRSIPQVDLI--ANARK-----GVRGYSHTVATENLDE-----FGIYKSTIF	455
Db	378	SY-EYDOLLRTDMIRCKNAKDFNIYPNVYNTSKCTITSASDSQTIKQESYSAYAQDAL	436
Qy	456	HPADGLSLIGGRGLGHYKIESGEGKTLHKASKTK---FTGYAGAVYDLNDNNSLYLSLQ	512
Db	437	YLTDNWITAVAGIRQYTYTAGKGRPENNVNTDSRDEQWTPKGLGVYKLPFSVSLFANYSQ	496
Qy	513	LYTPQTNLDAGKLLKPRQNGQFVGYKYSMDRLNARVSVFYMKDKDKNAAPLNPNNKK	572
Db	497	TFMPQSSIASYIGDLPPESPNAVYEGAKFLFDG-ITADIALFDIHKRNVLVYTESIGDET	555
Qy	573	TRYAALGKRYMEGVETRISGAVTPKWOIHAGYSYLHSOI-----KTASNSRDDGIFL	624
Db	556	IAKTA-GRVRSRGVEVDLAGALNTENINIIASYGYTDAKVLDEPDYACKPLPN-----	606
Qy	625	LMPKHSANLWTTYQV-----TPELTIGGGVNVMS--GITSSAGMHAGGYATFDMAAYRPF	677
Db	607	-VPRHTGSLFLTYDIIHNMGPNTLITFGGCHGVRSRATNGADYLLPGFYFADAFAYKQK	665

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:37:39 ; Search time 14.3071 Seconds
(without alignments)
2093.073 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776
Sequence: 1 MGQFMSVPRIMTATVLA.....MTFNIGSERTWNLRYSP 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	828.5	21.9	826	1	PBUA_PSESP
2	821.5	21.8	815	1	PFVA_PSEAE
3	767.5	20.3	819	1	PUPA_PSEPU
4	737.5	19.5	729	1	FHUE_ECOLI
5	715.5	18.9	720	1	PPTA_PSEAE
6	708.5	18.8	809	1	PUPB_PSEPU
7	470	12.4	760	1	YBIL_ECOLI
8	423	11.2	702	1	FOXA_SALTY
9	401.5	10.6	710	1	FOXA_YEREN
10	387.5	10.3	747	1	FHUA_ECOLI
11	374	9.9	735	1	FCF_ERMCH
12	303	8.0	673	1	FYUA_YEREN
13	300	7.9	673	1	FYUA_YERPE
14	267	7.1	663	1	CIRA_ECOLI
15	256.5	6.8	746	1	PFPA_PSEAE
16	221.5	5.9	345	1	YEGA_HAEIN
17	215.5	5.7	726	1	FATA_VIBAN
18	214	5.7	652	1	IRGA_VIBCH
19	214	5.7	700	1	YNCD_ECOLI
20	205	5.4	614	1	BTUB_SALTY
21	205	5.4	774	1	FECA_ECOLI
22	200	5.3	687	1	HENR_YEREN
23	198	5.2	732	1	LUTA_ECOLI
24	195	5.2	726	1	HMUR_YERPE
25	193.5	5.1	746	1	FEPA_ECOLI
26	193.5	5.1	746	1	RHTA_RHIME
27	191	5.0	614	1	BTUB_ECOLI
28	190.5	5.0	758	1	FCUA_YEREN
29	172	4.6	915	1	TBPI_NEIGO
30	167.5	4.4	908	1	TB12_NEIMB
31	165.5	4.4	725	1	HXC2_HAEIN
32	162.5	4.3	810	1	HPUB_NEIMA
33	161	4.3	911	1	TB11_NEIMB

34	158.5	4.2	723	1	Y262_HAEIN
35	157.5	4.2	810	1	HPUB_NEIMC
36	143	3.8	2003	1	YDBA_ECOLI
37	140.5	3.7	744	1	HXC1_HAEIN
38	137.5	3.6	1061	1	OAR_MYXXA
39	137	3.6	687	1	VYUA_VIBCH
40	137	3.6	2334	1	WAPA_BACCU
41	136	3.6	993	1	HGBC_HAEIN
42	130.5	3.5	4349	1	FAT2_HUMAN
43	129	3.4	699	1	CH11_BACCI
44	129	3.4	1046	1	HHUA_HAEIN
45	129	3.4	1462	1	GTFD_STRMO

ALIGNMENTS

RESULT 1	ID	PBUA_PSESP	STANDARD;	PRT;	826 AA.
AC	008017;				
DT	01-NOV-1995				
DT	01-NOV-1995				
DT	15-JUL-1999				
DE	Ferric-pseudobactin M14 precursor.				
GN	PBUA.				
OS	Pseudomonas sp. (strain M14).				
OC	Bacteria; Proteobacteria.				
OX	NCBI_TaxID=306;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=94104606; PubMed=8277948;				
RA	Morris J., Donnelly D.F., O'Neill E., McConnell F., O'Gara F.;				
RT	"Nucleotide sequence analysis and potential environmental				
RT	distribution of a ferric pseudobactin receptor gene of Pseudomonas				
RT	sp. strain M14."				
RL	Mol. Gen. Genet. 242:9-16(1994).				
CC	- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC				
CC	PSEUDOBACTIN M14.				
CC	- SURCELLULAR LOCATION: Outer membrane.				
CC	- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.				
CC	- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X73412; CAA51812.1; -				
DR	InterPro; IPR000531; TonB_boxC.				
DR	PFam; PF00593; TonB_boxC; 1.				
DR	PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.				
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.				
CC	Outer membrane; Iron transport; Transport; TonB box; signal; Receptor.				
FT	SIGNAL	1	44		
FT	CHAIN	45	826		
FT	SITE	110	119		
FT	SITE	809	826		
FT	SEQUENCE	826 AA;	90388 MW;	5D440D6AC7974CB CRC64;	

Query Match 21.9%; Score 828.5; DB 1; Length 826;

Best Local Similarity 29.9%; Pred. No. 2.8e-49; Indels 101; Gaps 26;

Matches 227; Conservative 129; Mismatches 303; Indels 101; Gaps 26;

16 TVLAALSSVFAQTADLETVHIIKGRSYNAIVTEKNGDYSSFAVTVGTRIPASLREIQ 75

115 TVLGPRATGAM-----ELAPTVNMSRL--GATTEGNSYTTGGVITIGGV-HSLKETPQ 166

76 SVSIIINQGVKRNVDTPQGLARKTPGLAVLSNDGRSSVYARGVYS-ENVIDGLP--- 131

167 SYVVMTRKMLDDQNTNTIQTQVMEKTPGIVYDPSPMGKRYFSGFRMSQYOYDGVLDI 226

```

QY 132 -----AQMOSINGTLENLFAFDREVRVGRGSLFDSGEMGGIVNLRKRPYKAFQGHAAA 187
DQ 227 GSSYVQADSFNS---DMAIYDREVLGAAGMMKGAGGTAGGVNFRKRGQDHTAHTQLSL 283
QY 188 GFTGTHQKQKAEADVSGSLNSDGSVRGR--VMAQT-----VCGASPRPAEKNNRHETFYAA 239
DQ 284 SAGTWNVRGQVDTGGPLNDSGTIGRAVVTETQYFYDVG-----SKDQIYYGA 335
QY 240 ADWDINPDTVLGAGLYQOORHLAP--YNGLPADANNKLPSLPQHVFGADMMKPKNSHDV 298
DQ 336 LDFDLSPDTTGLGPAWEDVDATPCWGLPRYADGSDLHLKSTCLNTAMNNQRKRTATY 395
QY 299 FADLKHFGNGGKGVKWMYSDRSDNSVAFAGSKLGMKTPAGRGCCNTADKACAVGLG 358
DQ 396 FADLKHQFNDDSLVAGVYSRNTQDMEYAFPSGAV----PVGATATNT-----LMLGSI 446
QY 359 TEIKOKALAFDASYSRPRLGNANTFEVFGADYNFRSTNEQGRITLVARGGL-----AL 413
DQ 447 YDQDQYGFDAVDGKDFADFCQOQHELTIGANASR-----SHKDDFYAVALPQRQNVL 500
QY 414 NEFRSIPQVD---LIANARKGVGRGSHYVATENLDEFGLYKSTFHPADGLSLIGGRIG 470
DQ 501 DNHHIPQDESYVLANASRGCPVDH-----IKQYGAYSIALKLADPLTLVLGSRVS 554
QY 471 HYKIES-----GEGKTLH-KASKT-KFTGYAGAVYDLNNDNSLYLSQLYTPO--TNL 520
DQ 555 WYKSDTDSVOYFRGEGTQVDTKSTGTQVTPPAGVLFDLNLDLTAYASVTDFTPGAYK 614
QY 521 DADGKLLPRQNGQEVGVKGYSDMDRLNARYSFFVRMKDKNAA-----APLNPNNK 572
DQ 615 TIDGTLPLVQSGVELGKGEWFGRLNSTPLNFTLQKDAQDDPCEDSSCSINSKG 674
QY 573 TRYAALGRKVRMGVETEISGATPQWQIHAGYSYLHSQIKTASNRDQDGLF--LMPKHS 630
DQ 675 VR-----AGCFRAEVSGEVIDRLQLLAGYTYTQTKVLEADATQDGVVNSVPRHL 726
QY 631 ANLWTTYQVTPEL---TIGGVNAMSQ---ITSAG---MHAGYATFDMAAYRTPKPK 682
DQ 727 LRWGDYLSGLPLDRVTIGAGVNAQTGNVRTSPIGGDNIDGAGYAVWNGRIGRYIDTWS 786
QY 683 LOINADNIFNRHYARVGANTFNIPGSERTWTANLRYSF 722
DQ 787 VALNGNNUFDKRYSTIGTGFNGFYGDPRNFVMSVKADF 826

RESULT 2
FPVA_PSEAE STANDARD; PRT; 815 AA.
AC FPVA_PSEAE Q51339;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferripyoverdine receptor precursor.
GN FPVA OR PA2396.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC NCBI_TaxID=287;
OX NCBI_TaxID=287;
RN [1]_TaxID=287;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-64.
RC STRAIN=CD10;
RX MEDLINE=93328663; PubMed=8335619;
RA Poole K., Neshat S., Krebs K., Heinrichs D.E.;
RT "Cloning and nucleotide sequence analysis of the ferripyoverdine
receptor gene fpva of Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:4597-4604 (1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RL Lamont I.L.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
[3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Raizel J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -1- FUNCTION: RECEPTOR FOR THE SIDEROPHORE, FERRIPYOVERDINE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY THE SIDEROPHORE, PYOVERDINE, AND UNDER IRON
STARVATION CONDITIONS.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-16 IS THE INITIATOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L10210; AAA25819.1; -.
CC EMBL; U07359; AAB60199.1; -.
CC EMBL; AE004666; AAG05786.1; -.
CC PIR; A40601; A40601.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC.1.
CC PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE_NEG.
CC PROSITE; PS01156; TONB DEPENDENT REC 2; 1.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
Complete proteome.
KW SIGNAL 1 43 POTENTIAL.
FT CHAIN 44 815 FERRIPYOVERDINE RECEPTOR.
FT SITE 798 815 TONB C-TERMINAL BOX.
FT CONFLICT 716 716 Y -> F (IN REF. 1).
FT CONFLICT 745 746 MISSING (IN REF. 1).
FT CONFLICT 750 750 S -> R (IN REF. 1).
SQ SEQUENCE 815 AA; 91167 MW; A11AB4AA5A290F35 CRC64;

Query Match 21.8%; Score 821.5; DB 1; Length 815;
Best Local Similarity 29.0%; Pred. No. 8.3e-49;
Matches 217; Conservative 145; Mismatches 312; Indels 75; Gaps 24;

QY 8 PRINTAATVLAALSSSVFAAQTADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGTIP 67
DQ 108 FQGNATISVAEAAADSSV-----DLGATMTISNQL--GTTEDSGSYTPGTIATRLV 159
QY 68 ASLREIPOSVSIITNQVKORNVDTFDQARKTPLGLRLVLSNDGRSSVYARGVESEYNI 127
DQ 160 LTPRETPOSITVTTRQNMDDFGLNIDVDMRHPTGITSAYDTRNNYARGFSLNNFYQ 219
QY 128 DGLPAQMQSI-----NGTLPLNLFAPDRVEVMRGPGLFDSGEMGGIVNLRKRPYKAFQ 183
DQ 220 DGIPTARNVGYSGAGNTLSDMAIYDRVEVLKAGTLLTGAGSLGATINLRKPTHEFKG 279
QY 184 HAAAGFGTHQYKABADVSGSLNSDGSVGRVMAQTGVASPRPAEKNNRHETFTYAAADWD 243
DQ 280 HVELGAGSDWNRSELDVSGPLTESGNVRGRAVAAAYQDKHSFMDHYERKTSVYTGILEFD 339
QY 244 INPDVLGAGLYQOORHL--APYNG-LPA--DANNKLPSLPQHVFGADMMKPKNSHDV 299
DQ 340 INPDMLTVGADYQNDPKSGWGSFPLFDSQGNRNDVSRFNNAGKWSWEQVTRTVF 399
QY 300 ADLKHFGNGGKGVKWMYSDRSDNSVAFAGSKLGMKTPAGRGCCNTADKACAVGLG 359
DQ 400 ANLEHNFANGWVGKVL---DHKINGYHAPLGAIMG-DWPA-----PDNSAKIVAQKY 448

```

QY 360 EIKOKLADASYRPPRLGNTANEFVIG--ADYNRPSTNEOGRTTLYARG-GLALNEF 416
 DB 449 TERTKSNSDIYLTGFGFGRHEHLVGTSAFSFHW-----EGKSYWMLRNNDNTTDF 503
 QY 417 RSLPQVDLLANARKGVRSHTVATENLDFGTYGKSTFHPADGLLIGGRIGHYKIES 476
 DB 504 -----INWDODIGKPRMGPSPQYIDKTRQGLGSGYMTARFNVTDLNLFLGGRVADV-7 557
 QY 477 GEGKTLHKASKTFGTGAGVYVDLNDNNSLYLSLSQLYTPQTN--LDADGKLKPRQNG 534
 DB 558 GUNPTIREG--RFIYVAGVYVDLNTYSVYASVYTDIFMWDQSWYDSSNKLLEPDEGQN 615
 QY 535 FEYGVGYSMTDRLNARVSFTRMKDYAAP-----LNPNNKTRYYAALG-KRMEGVET 568
 DB 616 YEIGIGELYLDGRLNTSLAYFEIHEENRAEDALVNSKPTNPATYYAKGIKAKTKGYEA 675
 QY 589 EISGATPCKMOIHAGSYLSHLSQIKTASNSRDDIIFILMPGHANLMTTVO--VTBELTI 645
 DB 676 EISGELAPGMQVAGYT--HKIIRDSGKK--VSTWEPDQSLSTYSYKFKGALDKLTV 730
 QY 646 GGGV-----NMSGITSSAGMHAGYATPDAMAAYFPTPKLQINADNIFNR 693
 DB 731 GGGARWQKSGWQVNVNPPSRWEK---FSGEDYVYVDLMARYQITDKLSASVAVNVVFDK 787
 QY 694 HYARVGGANTFNIPGSERTTANLRSF 722
 DB 788 TYTNIIGFYTSASY-GDPRLMFSTRWDF 815

RESULT 3

PUPA_PSEBU STANDARD; PRT; 819 AA.

AC P25184;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Ferric-pseudobactin 358 receptor precursor.
 GN PUPA.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxId=303;
 RN (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=WC358;
 RX MEDLINE=91260449; PubMed=1646376;
 RA Bitter W., Martug J.D., de Weger L.A., Tommaesen J., Weisbeek P.J.;
 RT "The ferric-pseudobactin receptor Pupa of Pseudomonas putida WC358:
 RT homology to TonB-dependent Escherichia coli receptors and specificity
 RT of the protein.";
 RL Mol. Microbiol. 5:647-655 (1991).
 CC - FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC
 CC PSEUDOBACTIN 358.
 CC - SUBCELLULAR LOCATION: Outer membrane.
 CC - SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X56605; CAA39942.1; -
 CC PIR: S15169; S15169.
 CC InterPro: IPR000531; TONB_boxC.
 CC Pfam: PF00593; TONB_boxC; 1.
 CC PROSITE: PS00430; TONB DEPENDENT REC 1; 1.
 CC PROSITE: PS01156; TONB DEPENDENT REC 2; 1.
 CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor.
 CC SIGNAL 1 47 POTENTIAL.
 CC CHAIN 48 819 FERRIC-PSEUDOBACTIN 358 RECEPTOR.

FT SITE 115 122 TONB BOX.
 FT SITE 802 819 TONB C-TERMINAL BOX.
 SQ SEQUENCE 819 AA; 91015 MW; 1B36164F86207951 CRC64;
 Query Match 20.3%; Score 767.5; DB 1; Length 819;
 Best Local Similarity 27.8%; Pred. No. 4,3e-45;
 Matches 214; Conservative 134; Mismatches 306; Indels 117; Gaps 26;

QY 8 FRINMTAAVTLAALSSVPAQAOTADLETHIKGOREYNALVTEKNQDYSFAVYVGTKIP 67
 DB 110 YQIDGNTVTVTASA-----AAKDQGLESATVNSAGLETTEBGTSYTRRTVSTATQNA 164
 QY 68 ASLREIPQSVSIIITNQVXDRNVDPDQARKTPGLRVLSDNG-RSSVYARGEYSEYN 126
 DB 165 LSIREFPQITIVTVTRQRMDDQHLGSWNEVLTQPGI-TWSQDGEFRFNYSRGSAINIYQ 223
 QY 127 IDGLPA-----QMSINGTIPNLPAFPRVAVMGPSGLFPSSGEMGIVLVLRKRPKARQ 182
 DB 224 PDGVTTYQDNQTRNMSTLMDVGLYDRIRIIRGATGLMTGAGDPSPAVVAVIRKRPREFK 283
 QY 183 GHAAAGFTHKQYKARADVSGSLNSDGYRGRVMAQTVGASPPAPAEKNR-----H 233
 DB 284 SHIQAGVSGMDYIRABADVSGPLTDDGRVGRFPA-----AKQDNHTFMWYTQOR 334
 QY 234 EEFYAADWDINDPVLGAGLYYQGRHLAPYNGLPA-----DANNKLPSLPQHYFVAGDWN 289
 DB 335 DVLGVVADVDDTVTARFGI--DRQYKVNAGAPGVPIIYTNQGPTRFSRSTSDARV- 390
 QY 290 KFRMNSHDVPAIDKHYFG-----NGYKRVGRVSRDRADSNVAFAGSKLGKTPAGRP 343
 DB 391 -----GYDDYTTNTYTFGEQQLAHDMQFPLAAVYVDVDRDPSFSSY-----YSTTNS 439
 QY 344 GGNLTADKACAVGLGTEIKOKALAPDASYRPPRLGNTANEFVIGADYNRPSTNEQGR 403
 DB 440 YLELDSTISAGIYV-AAQHQKVDATIGCPQLGCHLIVGNYLYEYKHA----- 493
 QY 404 TLYARGLALNEFRSLPOVDL-----IANARKG---VRGSHVATENLDFGIYGRS 453
 DB 494 ---RGD-----SGPDVININPYDMDNOTPKPGDDEIIRGIQYNSNR---QSGYFVAS 539
 QY 454 TRHPADGLSLIGGRIGHYKIE-----SGEKTLLHKASKTFGTGAGVYVDLNDNNSLY 507
 DB 540 RFLTDDTLHLILGARASNNRFPYALMRIGNEBAPYKRVBERGVTPAGVLYDILTMQSYV 599
 QY 508 LSLSLYTPQTNLDADGKLKPRQNGQFEVYKGSYMDRLNARVSFYMKKNAAAPLN 567
 DB 600 ASYTDLFFKQNNVDITGKPLDPEVGNVYELGMKGFELEGRLANALALYWKDNLAEISTN 659
 QY 568 ---PNNKTRYYAALGRVMEGVET-----EISGAVTPKQIHAGSYLSHLSQIKTASNSRD 619
 DB 660 EYVDPDSG-----GLASRAVDGAETKGVDELSEGVLPGMNVFTGYS--HTRTEDADGKR- 712
 QY 620 DGIFILMPGHANLMTTVOVTPD--LITGGGVNAMSGITSSAGMH-----AGYATPDA 671
 DB 713 --LTPQLPMDTRFNVNTYRLPGSEMERLITGGGVNMSKSTLNPARYNSHVTDQDYVTS 770
 QY 672 MAAYRTPKLIQINADNIFNRHYARVAGANTFNIPGSERTTANLRSF 722
 DB 771 MARYRNESLAATLVNINIFDKKTYA--QWAGSYGHGAPRNAVTVLRKDF 819

RESULT 4

FHUE_ECOLI STANDARD; PRT; 729 AA.

AC P16869; P77292;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE FhuE receptor precursor (Outer-membrane receptor for Fe(III)-coprogen,
 DE Fe(III)-ferrixanthine B and Fe(III)-rhodotulic acid).
 GN FHUE OR B1102.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90286919; PubMed=2162465;
RA Sauer U., Hantke K., Braun V.;
RT "Sequence of the fhuE outer-membrane receptor gene of Escherichia
colli K12 and properties of mutants."; Mol. Microbiol. 4:427-437(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Rau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."; Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
Sawai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map."; DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 37-50.
RC STRAIN=K12;
RX MEDLINE=87194585; PubMed=3032906;
RA Sauer M., Hantke K., Braun V.;
RT "Ferric-coprogen receptor FhuE of Escherichia coli: processing and
sequence common to all TonB-dependent outer membrane receptor
proteins."; J. Bacteriol. 169:2044-2049(1987).
CC -!- FUNCTION: REQUIRED FOR THE UPTAKE OF IRON(III) VIA COPROGEN,
FERRIOXAMINE B, AND RHODOTORULIC ACID.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
BE ACTIVE.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17615; CAA35616.1; -;
DR EMBL; AE000210; AAC74186.1; -;
DR EMBL; D90745; BAA35809.1; -;
DR EMBL; D90746; BAA35917.1; -;
DR PIR; S09262; S09262.
DR PIR; A26875; A26875.
DR EcoGene; EG10306; fhuE.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1;
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 36
FT CHAIN 37 729 FHUE RECEPTOR.
FT SITE 42 49 TONB BOX.

FT SITE 712 729 TONB C-TERMINAL BOX.
FT MUTAGEN 44 V->P: ABOLISHES TRANSPORT ACTIVITY.
FT MUTAGEN 46 V->P: ABOLISHES TRANSPORT ACTIVITY.
FT MUTAGEN 163 G->A,V,N,D,H: IMPAIRED TRANSPORT AND
GROWTH.
FT CONFLICT 363 G->C (IN REF. 1).
SQ SEQUENCE 729 AA; 81232 MW; 68ACEE7D110F76CD CRC64;
Query Match 19.5%; Score 737.5; DB 1; Length 729;
Best Local Similarity 27.3%; Pred. No. 4.2e-43;
Matches 211; Conservative 148; Mismatches 308; Indels 105; Gaps 26;
Qy 3 QFMSVFRINMTAATV-LAALSSSVFAAQTADLETVHVKQBSYNAIVTEKNGDVSSEFAVT 61
Db 11 QYOAITPSPLLAGCIALALLPSAFAAPATE-ETVIVEGSAT---APDGGNDYSVTSTS 66
Qy 62 VGTKIPASLRIPQSVSIITNQVKDRNVDTFDOLARKTGLRLVLSNDDGRSSVYARVE 121
Db 67 AGTKMQWTDIPQSVTIVSQORMEDQQLQTLGEVMENTLIGISQADSRLALYSRGFQ 126
Qy 122 YSEYNIDLPQMOS---INGTLPNLFADRVEMVRGPGSLFDSGEMGGIVNLRKPT 178
Db 127 IDNVMVDGIPYFESRNILGDSMALFERVEVVGATGLMTGTGNPSAAINVRKHAT 186
Qy 179 -KAFQGHAAAGFGTHKQYKABADVSGSLNSDGSVRGRVMAQTGCA---SPRAEKNNRHE 234
Db 187 SREFPKGVSABEYSGWNERVYVADLQSPLTEDGKIRARI---VGGYQNNDSWLDRYNSEK 242
Qy 235 TFYAA-ADWDINPDTVLGAGLYQORHL--APYNGLPA-DANNKLPSPHQHVFVGADNWK 290
Db 243 TFFSGIVDADLGDUTLLSAGYEYQRIQVNSTWGLFRWNTDSSNSYDRARSTAPDWAY 302
Qy 291 FKMNSHDVFDLKHVFGNGGKGVMRYSDRDADSNY----AFAGSKLGMKTP----- 339
Db 303 NDKEINKVEMTLKQFPADTQATLNATHSEVEFDSKNMYVDAYVKNADGMLVGPYSNYP 362
Qy 340 ---AGRPCCNTADDKACAVGLGTE-----IKQKALAFDASYSRPF--LGNTANEP-- 385
Db 363 GFDYVGGTGMNSGKRKYDADLDFADGSYELFGRQHNLMFGSGYSKQNNRYFSSWNIFFD 422
Qy 386 VIGADYNRFRSTNEQGRITLLYARGGLALNEPRSIPOVDLIANARKVGRGYSHTVATENLD 445
Db 423 EIGSFYN-FNG-----NFPQTD-----WSPQSLAQDDTT 450
Qy 446 EF-GIYKSTPHPADGLSLIGGGRGLGHYKESGEGKTLHKASKTKFTGYAGAVYDLNND 504
Db 451 HMKSLYAATRVTLADPLHLILGARYTNVRVDT---LTYSMKKNHTTPYAGLVFDINDNW 506
Qy 505 SLYLSQLYTPQTNLDADGKLLKPRQNGQEVGKSYMDRLNARVSFMRKDKNA-- 563
Db 507 STYASYTSIFQPNDRDSSGKYLAPITGNNYELGLKSDMNSRLTTLTALFRIEQDNVAQ 566
Qy 564 ---APLNPNNKKTVAALGKRVMGVEVETETSGVTPKQIHAGYS-YLHSQIKTKASNSRD 619
Db 567 STGPIPGSNGETAYKAVDGTVSKGVFEELNGALTDWQLTFGATRYI-----AEDNEG 620
Qy 620 DGIFLLMPKHSANLWTTYQ--VTPELTIGGVNAMSGITSSA-----GMHAGGYATFD 670
Db 621 NAVNPNLPTTKMFTSYRLPVMPPELTVGGVNNQNRVYTDVTPYGTFRAEQGSYALVD 680
Qy 671 AMAAYRTPKLLQINADNINRHHYARVGGANTFNIPGSERTWTANLRYSF 722
Db 681 LFTRYQVTKNFSLOQNVNPLFDKTYDITNVEGSIYV---GTPRNFSGITGYOF 729
RESULT 5
FPTA_PSEAE
ID FPTA_PSEAE STANDARD; PRT; 720 AA.
AC P42512;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fe(III)-pyochelin receptor precursor.

GN PPTA OR PA4221.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-51.
 RC STRAIN=PAO / 1A602;
 RX MEDLINE=94117363; PubMed=8288523;
 RA Ankenbauer R.G., Quan H.N.;
 RT "FpA, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a phenolate siderophore receptor homologous to hydroxamate siderophore receptors."
 RT J. Bacteriol. 176:307-319 (1994).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAO1;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brady L.L., Coulter S.N., Folger K.R., Kas A., Lathig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen."
 RT Nature 406:959-964 (2000).
 RT Nature 406:959-964 (2000).
 CC -!- FUNCTION: HIGH-AFFINITY OUTER MEMBRANE RECEPTOR REQUIRED FOR THE TRANSPORT OF Fe(III)-PYOCHELIN.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, U03161; AAC43213.1; -;
 DR EMBL, AB004839; AAG07609.1; -;
 DR InterPro, IPR000531; TONB_boxC.
 DR Pfam, PF00593; TONB_boxC.1.
 DR PROSITE, PS00430; TONB-DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE, PS01156; TONB-DEPENDENT_REC_2; 1.
 KM Outer membrane; Iron transport; Transport; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 38
 FT CHAIN 39 720
 FT SITE 703 720 TONB C-TERMINAL BOX.
 FT SEQUENCE 720 AA; 79992 MW; DA796313116B0CC2 CR664;
 Query Match 18.9%; Score 715.5; DB 1; Length 720;
 Best Local Similarity 27.2%; Pred. No. 1.3e-41;
 Matches 200; Conservative 137; Mismatches 323; Indels 75; Gaps 20;
 QY 21 LSSVRAAQTAD-----LFTVHIKQGRVNAIVTEKNKGDSSRAVYVGTIKIPSLAR 72
 DB 28 LALSPILAAVADARKDGETELPDMVSGE--STSAIQPRG-----VTTLGAVPLKPRF 78
 QY 73 IPQSVAIITNOQVKNVDVTFDOLARKTPGLRIYLSNDGSSVYARGVSEYSEINIGLPA 132
 DB 79 LQGSASVIERHERLEQONTLSLDEAMQACVYQPFQQLTAYVAGFKYDSLELDGVR 138
 QY 133 QMOSINGTLPNLFAPRVEVMRPGSGLPSSGEMGIVLVIRKRPKTAQGHAAAGFTH 192
 DB 139 LIGNTASSQDMVAIVERVELIRSGNGLHGTGNPAATVNLVRKRPREFAASTTLISGRV 198
 QY 193 KQYKARADVSGSLNSDGSVGRMAQTGASPRP-----AEKNRHETFYAADMDINPD 247
 DB 199 DRIAEVDVGGPGLSASGNVGRRAVA---AYEDRDYVDVADQSTR--LITGVTEFLDSP 253

QY 248 TVLGAGYLTYQ--RHLPVINGLPADANNKLPSPLOHVFVADNNKFRMNSHDVPAIDKHYF 306
 DB 254 TLITVGAQYQHDISTLNNMAGVPMADGSLGSRDYLVDWDFRFXMDTYRAFGSLEQOL 313
 QY 307 GNGGYKGMKRSDDRDADSNVAFAQSKLGMKTPAPRPGCNTADDKACANGLGEIYQKAL 366
 DB 314 GGGWKGKVAEY--DEADSLRYAGSPGAIIDPOTG-----DGGQLGAAVYKFSIOR 363
 QY 367 APDASYSRPFRLGNTFANEVIGADYRPFSTNEOGRTTLTYARGLALNEFRSIPQ--VDL 424
 DB 364 SLDANINGVRLPGLTHELIGGVTAQGETRODTAR-----FLNLPNTPVAV 410
 QY 425 IANARKVR---GSHVATENLDEFGIYKSTFHPADGLSLGGGRIGHYKIESBECK 480
 DB 411 YRWDPHGVPRPQIGQYTSPTGTTTQKGLYALGRILAEPLTLVVGRESW--DQDTPA 468
 QY 481 TLHKASKTFPGYAGAVYDLNDDNSLYLSQLYTPQNTLDA--DGTLKPRGNGQREYV 539
 DB 469 TRPKPR--QFTPYGGLIWPARDMSVYSAVYQPOADQRTWNSPPLSPVSGKTYETGI 527
 QY 540 KGSYMDRLNARVSFYRMKNA--PLPNKKTRVYALGKRVMEGVETEISGAV 594
 DB 528 KGEIADGRNLSLAARIDLENNPQEDPHRPPNN--PFYISGKVRSGFELBGTGL 585
 QY 595 TPQMQIHAGSYLHSQ-IKTASNSRDDGIFLLMPKHSANLWTTYQV--TPELLTGGVN 650
 DB 586 TPYMSLSAGYVTSFTYLDQSDQSTGRVSTFPLRLILRLMSYDLFPQDRRWSVGGGLQ 645
 QY 651 ANSGIT---SSAGMAGGYATPDMAAAYFTFKLQINADINFNHRYARVAGANTFNI 707
 DB 646 AOSDYSVDYRGVSMRGGGALVNMRLGYKIDHWTAAVVNNLFDRTYQSLSPNMNMR 705
 QY 708 PSERTTANLRYSF 722
 DB 706 YGPRRPNVSLGAP 720
 RESULT 6
 PUBS_PSEPU
 ID PUBS_PSEPU STANDARD; PRT; 809 AA.
 AC P38047;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Ferric-pseudobactin Bm7/Bn8 receptor precursor.
 GN PUBS.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WCS358;
 RX MEDLINE=93116856; PubMed=8392140;
 RA Koster M., van de Vosseberg J., Jeong J., Weisbeek P.J.;
 RT "Identification and characterization of the pubB gene encoding an inducible ferric-pseudobactin receptor of Pseudomonas putida WCS358."
 RT Mol. Microbiol. 8:591-601 (1993).
 RL -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORES FERRIC PSEUDOBACTIN Bn8 AND FERRIC PSEUDOBACTIN Bm7, IRON CHELATING MOLECULES THAT ALLOW THE ORGANISM TO EXTRACT IRON FROM THE ENVIRONMENT, SPECIALLY UNDER IRON-RESTRICTED CONDITIONS.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: BY IRON LIMITATION IN ADDITION TO THE PRESENCE OF ONE OF THE TWO COGNATE PSEUDOBACTINS Bn8 OR Bm7.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

[illegible]


```
QY 173 VRKRPTKAFQGHAAAGFGTHKQYKAEADYSGSLNSDGSVGRVMAQTVCASPRPAEKNR 232
D 166 TSKPAPFADAGGSKVLKPAAGNNNGAFAFDVTGLDNERVAABLSCGMYRVSQFTPLKEE 225
QY 233 HETFEVAAAADWDNPDTVLG-----AGLYQORHLAPVNGLPADANNKPLSL 278
D 226 RYALMPSLWTRITDRTRLDLMAYPHRDPGGSHSLPYQGT-VVPYNG-----GKI 275
QY 279 PQHVFFVGA-DWNKFKMNSHDVADLKHYPGNGYGVKGMRYSDRDADSNYAFAGSKLGMK 337
D 276 SNTFFEGEDDYKYDRRENMGVNIHFLPDNGWSVRQKRLYLTKVTLNOVVAAGWLN-- 333
QY 338 TPAGRPCNTADDKACAVGLGTEIKQALAPDASVSRPRLGNTANEFVIGADYNRFRST 397
D 334 -----ETALNRGYS-----GSGEKMSAIALDNLQDGSVDTGAINHRLVLDYDQ----- 377
QY 398 NEQGRITLVARGGLALNEFRSIPQVDLIANARKVRGYSHTVATE--NLDFEFGIYKSTF 455
D 378 DRSNHTTG-----GAPPPDAFNPVYGAQPDYITLYSREKHKLRQTYGLQDM 428
QY 456 HPADGLSLTGGRLGHYKIESGEGKTLHKA-----SKTKFTGYAGAVYDLNDNNSLYLSLS 511
D 429 -SWDRWFTLGGRYDRVSVSNID--KLHDSRSLDKNNVSTRAALLYLFDNGVAPYLSVS 485
QY 512 QLYTPTNLDADGKLLKPRQGNQFVGYKGSYMDRLNARVSFYRKQKNAAPLNPNK 571
D 486 TAPFTSTFADENGVLNLEPMKQWEAGVYKPEPGGNSQFSAAVYRINQTNATKEPTDP 545
QY 572 KTRYAALGRVMEVETEISGAVTPKQWQHAGYSYHLHSQIKTASNRDDGIPLLMPKESA 631
D 546 ----YRSIGIESGKELEAISLHLSVRLOQAAYTYDIRKKSSPOEGKRAVYAPRQA 602
QY 632 NLWTYTVQ-----TPELTIGGVNAMSGITSS-AGMHA-GGYATFDAMAAYRTP----KL 681
D 603 SALSVDKSLGLEGLTGLSGIRYNGVTSRLNTHLTPSYLVDVMVGYDLSLGLNGL 662
QY 682 KLIQINADNIFNRHYARVGANTFIPGSERTWTANLRYSP 722
D 663 SAQLNVNLTDRKYVAACNSL-SYCVFGAERSIVGVSNAF 702

RESULT 9
FOX_A_YEREN STANDARD; PRT; 710 AA.
AC Q01674;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ferrioxamine receptor precursor.
GN FOXA.
OS Versinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_taxid=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RX MEDLINE=92349959; PubMed=1640832;
RA Baemler A.J., Hantke K.
RT "Ferrioxamine uptake in Versinia enterocolitica: characterization of
RL the receptor protein FOXA.";
RL Mol. Microbiol. 6:1309-1321(1992).
RN [2]
RP REVISIONS.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA Baemler A.J.;
RL Submitted (Nov-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: FERROXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE
CC TONB PROTEIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60447; CAA42975.1; -.
DR PIR; S22673; S22673.
DR HSP; P06971; IQQJ.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW Transport; TonB box.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 710 FERROXAMINE RECEPTOR.
FT SITE 28 35 TONB BOX.
FT SITE 693 710 TONB C-TERMINAL BOX.
FT TRANSMEM 29 37 POTENTIAL.
FT TRANSMEM 65 73 POTENTIAL.
FT TRANSMEM 91 99 POTENTIAL.
FT TRANSMEM 106 114 POTENTIAL.
FT TRANSMEM 137 145 POTENTIAL.
FT TRANSMEM 152 160 POTENTIAL.
FT TRANSMEM 180 188 POTENTIAL.
FT TRANSMEM 194 202 POTENTIAL.
FT TRANSMEM 208 216 POTENTIAL.
FT TRANSMEM 259 267 POTENTIAL.
FT TRANSMEM 271 279 POTENTIAL.
FT TRANSMEM 293 301 POTENTIAL.
FT TRANSMEM 309 317 POTENTIAL.
FT TRANSMEM 353 361 POTENTIAL.
FT TRANSMEM 370 378 POTENTIAL.
FT TRANSMEM 427 435 POTENTIAL.
FT TRANSMEM 443 451 POTENTIAL.
FT TRANSMEM 476 484 POTENTIAL.
FT TRANSMEM 491 499 POTENTIAL.
FT TRANSMEM 517 525 POTENTIAL.
FT TRANSMEM 531 539 POTENTIAL.
FT TRANSMEM 555 563 POTENTIAL.
FT TRANSMEM 567 575 POTENTIAL.
FT TRANSMEM 579 587 POTENTIAL.
FT TRANSMEM 610 618 POTENTIAL.
FT TRANSMEM 624 632 POTENTIAL.
FT TRANSMEM 649 657 POTENTIAL.
FT TRANSMEM 671 679 POTENTIAL.
FT TRANSMEM 684 692 POTENTIAL.
FT TRANSMEM 702 710 POTENTIAL.
SQ SEQUENCE 710 AA; 78382 MW; 82EB6EC1546900B8 CRC64;

Query Match 10.6%; Score 401.5; DB 1; Length 710;
Best Local Similarity 23.3%; Pred. No. 5e-20;
Matches 182; Conservative 127; Mismatches 334; Indels 139; Gaps 30;

QY 6 SVFRINMTAATVLAALSSSVFAAQADLETVHKQRSYNAIVTEKNGDYSSFAVTYGTGK 65
D 3 SAFIIRKSA--ILCCLAMFIPLASIAD-DTIEVTAKAGHEADL--PTSGYTATTTKGATK 57
QY 66 IPASLRIPQSVSIITNQVKDRNVDTFDQLARKTPGLRVLSNDDGRSSVY----ARGYE 121
D 58 TDQPLILTAAQSVSVVTRQMDQDVATVNOALNYPG--VFTGFSGGATRYDTVALRGFH 115
QY 122 YSEYN---IDGLPAQMOSINGTLLENL---FAPDRVEMRGPSGLFDSSGEMGGIVNLVR 174
D 116 GGVNNTFDGL--RLISDGSYVNVLDVFWFLERIDVIGKPSALLYQOSIPGGVVMWTS 173
QY 175 KRPTKAFQGHAAAGFGTHKQYKAEADYSGSLNSDGSVGRVMAQTVCASPRPAEK--NNR 232
D 174 KRPOFTSEGHFRLLTAGNNITQVAADFVTDALSEHWAFR-----LTGITRNSDTMVDHQ 226
QY 233 HETFYAAD---WDINPDTVL-----GAGLYQORHLAPVNGLPADANNKPLSLP 279
```

```

Db      227 REERRVIAISLLMQPDENTSLLRANLQKDPSSGKY-----HSNVPADGSIYGGKLS 277
Qy      280 QHFFVADWNKFKFGNSHDVPAFLKHFFGNGGYGKGMRYSDRSDADSVAAGSKLGKTP 339
Db      278 RGFDE-----SHNVFKRQOITYSEFSKFPDDWSFRQ-NMSYTHSNLTQLEVTQ 329
Qy      340 AGRPGCNTADKACAVGLGTEIKKALAPDASYSRPRGLTANTANEPVIGADYNRFRS--- 396
Db      330 GGNNSDRTLNNRYYS---GEDSLNFAVDNQLADRLTAAVKHVLLGVDPQKFRNNLR 386
Qy      397 -----TNEGRTTLVARGGLALNEFRSIPQVDLIANARKGVRSHTVATENL 444
Db      387 SDSAVATPLNPPYGVSGSSTLYS-----DYLRT-----TPGINTSYLSRRY 427
Qy      445 DERGIYKSTFFHPADGLSLIGGRGLGHYKIE-----SGEGTKLKASTKFTGYAGAV 497
Db      428 EGGGVTLQDEM-TLDWMHNLTSGRYDRMKTENINNTANSTDEKTDNHA-----GRASL 481
Qy      498 YDLNNDNSLYLSQLYTPQTNLDADGKLKPRQNGFEVQ-YKGSYMDRLNARVSFY 555
Db      482 YSPDSGISPYVSQAITPBLPDAQOKLLKPMTSSEQVEGIIYQPGGSTSLVSALYDL 541
Qy      556 RAKD-KVAAAPLPPNKKTRYALGKRVMEGVETEISGAVTPKKQIHAGYSYLSQIKTA 614
Db      542 TQNDVANRAVP-----ATYYVPAGVNSQGLHLEARSQISDRLSVLAGTYNRVFXKA 595
Qy      615 SNSRDGIFFLMPKHSANLMTTVOVPELITGGVNAMSGI-----TSSAGMHAGY 666
Db      596 IDNDGNTVLABSNWASMLAOYEA-----GYGVNGAGIRYIQOMADANTLVPST 649
Qy      667 AAFDAAAY---RFTPKLK---LQINADNIFNRHYARVAGANTFNI PGSERTWTANLRY 720
Db      650 TEGDASVRADLGWMAASLKGAFQVLVNNINADKKVYAAC-YSTSYCWAGRESRVQTVGY 708
Qy      721 SF 722
Db      709 DF 710

RESULT 10
FHUA_ECOLI
ID FHUA_ECOLI STANDARD; PRT; 747 AA.
AC P06971; P75665; P71280;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferrichrome-iron receptor precursor (ferric hydroxamate uptake)
DS FHUA OR TONA OR B0150.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CX NCBI_TaxId=562;
OK (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-47.
RP STRAIN=K12;
RX MEDLINE=86085668; PubMed=3079747;
RA Coulton J.W., Mason P., Cameron D.R., Carmel G., Jean R., Rode H.N.;
RT "Protein fusions of beta-galactosidase to the ferrichrome-iron
RL receptor of Escherichia coli K-12."
RL J. Bacteriol. 165:181-192(1986).
RP (2)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RL the 2.4-4.1 min (110,917-193,643 bp) region."
RL Nucleic Acids Res. 22:1637-1639(1994).
RP (3)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;

```

```

RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RP (4)
RP SEQUENCE OF 482-647 FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Laskari D., Lew H., Lin D., Namach A., Oelner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
RP (5)
RP SEQUENCE OF 723-747 FROM N.A.
RX MEDLINE=88038363; PubMed=2823072;
RA Burkhardt R., Braun V.;
RT "Nucleotide sequence of the fhuc and fhud genes involved in iron
RT (III) hydroxamate transport: domains in Fhuc homologous to
RT ATP-binding proteins."
RL Mol. Gen. Genet. 209:49-55(1987).
RP (6)
RP REVIEW.
RX MEDLINE=94265918; PubMed=7515827;
RA Braun V., Killman H., Benz R.;
RT "Energy-coupled transport through the outer membrane of Escherichia
RT coli small deletions in the gating loop convert the fhua transport
RT protein into a diffusion channel."
RL FEBS Lett. 346:59-64(1994).
RP (7)
RP ION CHANNEL.
RX MEDLINE=96203106; PubMed=8617231;
RA Bonhivers M., Ghazi A., Boulanger P., Letellier L.;
RT "Fhua, a transporter of the Escherichia coli outer membrane, is
RT converted into a channel upon binding of bacteriophage T5."
RL EMBO J. 15:1850-1856(1996).
RP (8)
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 34-747.
RX MEDLINE=99074366; PubMed=9856937;
RA Ferguson A.D., Hofmann E., Coulton J.W., Diederichs K., Welte W.;
RT "Siderophore-mediated iron transport: crystal structure of Fhua with
RT bound lipopolysaccharide."
RL Science 282:2215-2220(1998).
RP (9)
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 53-747.
RX MEDLINE=99081293; PubMed=985695;
RA Locher K.P., Rees B., Koebnik R., Mitschler A., Moulinier L.,
RA Rosenbusch J.P., Moras D.;
RT "Transmembrane signaling across the ligand-gated Fhua receptor:
RT crystal structures of free and ferrichrome-bound states reveal
RT allosteric changes."
RL Cell 95:771-778(1998).
RP (10)
RP X-RAY CRYSTALLOGRAPHY (2.95 ANGSTROMS) OF 34-747.
RX MEDLINE=20306978; PubMed=10850805;
RA Ferguson A.D., Braun V., Fiedler H.-P., Coulton J.W., Diederichs K.,
RA Welte W.;
RT "Crystal structure of the antibiotic albomycin in complex with the
RT outer membrane transporter Fhua."
RL Protein Sci. 9:956-963(2000).
RP (11)
RP FUNCTION: THIS RECEPTOR BINDS THE FERRICHOME-IRON LIGAND. IT
RP INTERACTS WITH THE TONB PROTEIN, WHICH IS RESPONSIBLE FOR ENERGY
RP COUPLING OF THE FERRICHOME-PROMOTED IRON TRANSPORT SYSTEM. ACTS
RP AS A RECEPTOR FOR BACTERIOPHAGE T5 AS WELL AS T1, PH180 AND
RP COLICIN M. BINDING OF T5 TRIGGERS THE OPENING OF A HIGH
RP CONDUCTANCE ION CHANNEL. CAN ALSO TRANSPORT THE ANTIBIOTIC
RP ALBOMYCIN.
RP SUBUNIT: MONOMER.
RP -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
RP -I- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
RP -----
RP This SWISS-PROT entry is copyright. It is produced through a collaboration

```

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M12486; AAB61768.1; -;
DR EMBL; D26562; BAA05598.1; -;
DR EMBL; AE000124; AAC73261.1; -;
DR EMBL; U70214; AAB08580.1; -;
DR EMBL; X05810; CAA29253.1; -;
DR PIR; A25196; QRECFE.
DR PIR; S06358; S06358.
DR PIR; S45219; S45219.
DR PDB; 1BY3; 13-JAN-99.
DR PDB; 1BY5; 13-JAN-99.
DR PDB; 1FCP; 13-JAN-99.
DR PDB; 2FCP; 13-JAN-99.
DR PDB; 1OQJ; 05-JUN-00.
DR PDB; 1OQC; 05-JUN-00.
DR EcoGene; EG10302; fluA.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transpor; TonB box; Signal;
KW Phage recognition; Receptor; Transmembrane; 3D-structure;
KW Complete proteome.
FT SIGNAL 1 33
FT CHAIN 34 747
FT DOMAIN 34 747 FERRICHRONE-IRON RECEPTOR.
FT TRANSMEM 193 201 PERIPLASMIC.
FT DOMAIN 202 206 EXTRACELLULAR.
FT TRANSMEM 207 215 PERIPLASMIC.
FT DOMAIN 216 222 EXTRACELLULAR.
FT TRANSMEM 223 231 PERIPLASMIC.
FT DOMAIN 232 245 EXTRACELLULAR.
FT TRANSMEM 246 255 PERIPLASMIC.
FT DOMAIN 256 259 EXTRACELLULAR.
FT TRANSMEM 260 268 PERIPLASMIC.
FT DOMAIN 269 312 EXTRACELLULAR.
FT TRANSMEM 313 321 PERIPLASMIC.
FT DOMAIN 322 326 PERIPLASMIC.
FT TRANSMEM 327 335 EXTRACELLULAR.
FT DOMAIN 336 387 EXTRACELLULAR.
FT TRANSMEM 388 396 PERIPLASMIC.
FT DOMAIN 397 404 EXTRACELLULAR.
FT TRANSMEM 405 413 EXTRACELLULAR.
FT DOMAIN 414 464 PERIPLASMIC.
FT TRANSMEM 465 473 EXTRACELLULAR.
FT DOMAIN 474 477 PERIPLASMIC.
FT TRANSMEM 478 486 EXTRACELLULAR.
FT DOMAIN 487 508 EXTRACELLULAR.
FT TRANSMEM 509 517 PERIPLASMIC.
FT DOMAIN 518 522 EXTRACELLULAR.
FT TRANSMEM 523 531 PERIPLASMIC.
FT DOMAIN 532 551 EXTRACELLULAR.
FT TRANSMEM 552 560 PERIPLASMIC.
FT DOMAIN 561 565 EXTRACELLULAR.
FT TRANSMEM 566 574 EXTRACELLULAR.
FT DOMAIN 575 601 PERIPLASMIC.
FT TRANSMEM 602 610 EXTRACELLULAR.
FT DOMAIN 611 613 PERIPLASMIC.
FT TRANSMEM 614 622 EXTRACELLULAR.
FT DOMAIN 623 645 PERIPLASMIC.
FT TRANSMEM 646 654 EXTRACELLULAR.
FT DOMAIN 655 661 PERIPLASMIC.
FT TRANSMEM 662 670 EXTRACELLULAR.
FT DOMAIN 671 689 PERIPLASMIC.
FT TRANSMEM 690 698 PERIPLASMIC.
FT DOMAIN 699 705 PERIPLASMIC.
FT TRANSMEM 706 714

FT DOMAIN 715 737
FT TRANSEM 738 746
FT DOMAIN 747 747 PERIPLASMIC.
FT SITE 40 47 TONB BOX.
FT SITE 730 747 TONB C-TERMINAL BOX.
FT BINDING 114 114 FERRICHRONE.
FT BINDING 133 133 FERRICHRONE.
FT BINDING 148 149 FERRICHRONE.
FT BINDING 277 277 FERRICHRONE.
FT BINDING 279 279 FERRICHRONE.
FT BINDING 346 346 FERRICHRONE.
FT BINDING 348 348 FERRICHRONE.
FT BINDING 424 424 FERRICHRONE.
FT BINDING 735 735 FERRICHRONE.
FT DISULFID 351 362
FT DISULFID 725 731
FT CONFLICT 609 610
FT CONFLICT 737 737
SQ SEQUENCE 747 AA; 82182 MW; 1C2B251D1990E444 CRC64;
Query Match 10.3%; Score 387.5; DB 1; Length 747;
Best Local Similarity 22.0%; Pred. No. 4.9e-19;
Matches 175; Conservative 122; Mismatches 362; Indels 135; Gaps 30;
Qy 6 SVFRINMTAATVLAALSSSVFAAQADLETVHIKQSRNAIVTEKNGDYSSFAVTVGTK 65
Db 12 SLRKIAVVAVTAVGSMVYAQAQVEPKEDITVTAAAPQESAWGPAATIAAROSATGK 71
Qy 66 IPASLRIPQSVSIITNQVKDRNVDFDQLARKTPGLRVLSDDDGRSSVY----- 116
Db 72 TDTPIQKVPQISVVTAEEMALHPKSVKEALSYPGVSV--GTRGASNTVDHLIIRGFA 129
Qy 117 ARGYEYSEYNTIDGLPAQMSINGTLPNLPAPDRVEMRGPSGLFDSSGEMGGINLVKVR 176
Db 130 AEGSQNNY-LNGLKQGNFYNDVIDPYMLERAIEIMRGPSVLYLKGSSPGGLLNVMVSKR 188
Qy 177 PT----KAFQGHAAAGFTGHQYKAEADVSGSLNSDG--SVGRVMAQTVGASPRPAEK 230
Db 189 FTTEPLKEVQPKA----GTDLSLFTQTFDFDSLDDDDGVSYRLTGLARSANAQ----QKG 240
Qy 231 NRHETFAAADWDINPD-----TVLGAGLYQORHLAPYNGLPADAN-NKLPS---LPQH 281
Db 241 SEQRYAIAPAFTWRPDDKTNFTFL--SYFONEPETGYGWLPEKGVETPLPNGKRLPTD 298
Qy 282 VFGADWNKFKQNSHDVPADLKHVFGNGYGVKGVMRYSDRDADSNYPAGSKLGMKTPAG 341
Db 299 FNEGAKNNTYSRNEKVMGVSPDHEFND-----TFTVR---QNLRAFNKNTSQNSVYG 347
Qy 342 RPGCN--TADDKACAV-----GLGTEIKOKALAFDASYSRPFRLGNTANEF 385
Db 348 YGVCSDPANAYSQCAALAPADKGHYLARKYVVDDEKLQNFSDVTQLQSKFATGDIDHTL 407
Qy 386 VIGADYNFRSTNEQRTTLYARGGLALNEPRSIPOVDLIANARKVGRVGYSHTVATENLD 445
Db 408 LTGYDFWEMR--NDINAWFGYDDSVPLLNLYNPV-NTDFDNADKPANSGPYRLNKQ-K 463
Qy 446 EFGIY-----GKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKF 490
Db 464 QTGVYVQDAQQWDKVLVTLGGRYDWAQDES LNRVAG-----TTDKRDDKQF 509
Qy 491 TGYAGAVYDLNDNNSLYLSLSQLYTPQTNLDADGKLLKPROGNQFVGYKGSYMDRLNA 550
Db 510 TWRGVNYLFDNGVTPTPYFSYSESPSSQVQKQGNIFAPSKGQYGVKVPDRPLV 569
Qy 551 RVSPYRMKDKNAAAAPLNPNNKKTTRYAALGRV-MEGVETEISGAVTPKWQIHAGYSYLHS 609
Db 570 TCAVYNTKTN--NLMADPGSFFSVEGGEIRARGVEIEAKAALSASVNVVGSYTYTDA 626
Qy 610 QIKTASNRDDGIFLLMPKHSANLWTVYQVTPPELTIGGVNAMSIGTSSAG----- 660
Db 627 EYTTDTTYKGN-TPAQVPKHMASLWADY-----TFDDG--PLSGLTLGTGRTVGSYSG 677
Qy 661 -----MHAGGYATFDMAAYRFT-----PKLKLQINADNIFNRHYARVGGANTFN-IP 708

Db	678	DPANFKGYSYIVDALRYDLARVGMSNVALHVN--NLEDFREYVASC--FNTRYGCEW	733
Oy	709	GSERTWTANLRYSF 722	
Db	734	GAERQVAVATATFRF 747	
RESULT 11			
FT_ERMCH	STANDARD;	PRT, 735 AA.	
AC	Q47162;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	FerriChrysoactin receptor precursor.		
GN	FT.		
OS	Erwinia chrysanthemi.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Pectobacterium.		
OX	NCBI_TaxID=556;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=3937;		
RC	MEDLINE=96165286; PubMed=8576065;		
RX	Savage C., Franta T., Expert D.;		
RA	"Analysis of the Erwinia chrysanthemi ferriChrysoactin receptor		
RT	gene: resemblance to the Escherichia coli fepA-fes bidirectional		
RT	promoter region and homology with hydroxamate receptors.";		
RL	J. Bacteriol. 178:1227-1231(1996).		
CC	-1- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING		
CC	CHRYSOACTIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS THE		
CC	BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.		
CC	-1- SUBCELLULAR LOCATION: Outer membrane.		
CC	-1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	CC between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	CC the European Bioinformatics Institute. There are no restrictions on its		
CC	CC use by non-profit institutions as long as its content is in no way		
CC	CC modified and this statement is not removed. Usage by and for commercial		
CC	CC entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X87967; CAA61205.1; -.		
DR	HSSP; P06971; 2PCP.		
DR	InterPro; IPR00531; TonB boxC.		
DR	Pfam; PF00593; TonB_boxC; 1.		
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.		
DR	PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.		
KW	Signal; Receptor; Transmembrane; Iron transport;		
KW	Transport; TonB box.		
FT	SIGNAL	1 38	POTENTIAL.
FT	CHAIN	39 735	FERRICCHRYSOACTIN RECEPTOR.
FT	SITE	45 52	TONB BOX.
FT	SITE	718 735	TONB C-TERMINAL BOX.
FT	DOMAIN	39 45	PERIPLASMIC (POTENTIAL).
FT	TRANSMEM	46 54	POTENTIAL.
FT	TRANSMEM	55 82	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	83 91	POTENTIAL.
FT	DOMAIN	92 107	PERIPLASMIC (POTENTIAL).
FT	TRANSMEM	108 116	POTENTIAL.
FT	DOMAIN	117 124	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	125 133	POTENTIAL.
FT	DOMAIN	134 169	PERIPLASMIC (POTENTIAL).
FT	TRANSMEM	170 178	POTENTIAL.
FT	DOMAIN	179 179	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	180 188	POTENTIAL.
FT	DOMAIN	189 194	PERIPLASMIC (POTENTIAL).
FT	TRANSMEM	195 203	POTENTIAL.
FT	DOMAIN	204 222	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	223 231	POTENTIAL.
FT	DOMAIN	232 261	PERIPLASMIC (POTENTIAL).

Query Match	Beet Local Similarity	Matches 184; Conservative 119; Mismatches 341; Indels 130; Gaps 31
14 AATVLAALSSVFAAQTADLETVHKGGRS-----NVAIVTEKNGKDDVSFAVTGFKIP 67	9.9%; Score 374; DB 1; Length 735;	
27 ASTLMAAHQAQ--AAETTCADMTIVSANAGESVTAPLKGIVAKESAS-----GTKTS 76	23.8%; Pred. No. 4.1e-18;	
68 ASLRPEIPOSVEIITNOQYKDRNVDTPDOLARKTPELRYLSNDGRRS-----VYAR-GVEY 122		
77 TPLIKTPQSVTVVATATKMDAQAVSSVSHALANYSG--VVTNRRGSSNNRDEVIARPRVY 134		
123 SEYNIDGLPAQMOSINGTL--PNLEAFDRVEVYMRGSGLPDSSGEMGGIVNLVVRKPTYA 180		
135 APKFIDGLSTYSGSGSGSTIGKKNPWLIRERVEVHGPASVLYQGVNPGGLISMTSKRPTA- 193		
181 FQGHAAAFGTHKQYKAA--DVSGSLNSDGSVRRVRAVQAQTVGASAPRAPKNNRHETFFA 238		
194 -ETTRKQVFSAGNQLGPAAPDFGALLNDKTLTYRL--DGLASTKHEFVVKSKORRIAV 250		
239 AAD--WDINPTVLGAGLYQOORHLAPVNGLPADANNKLPSPQGVFVGA----- 286		
251 APSLTWLENPDTSTFLTLYSYQNDPRAAGI-----RNLPLKGIYVEASAGIYIPDLNVS 303		
287 --DNWKFKNSHDVFADIKHYFGANGGYKVGKMRYSDDRADSNYAFAGSLGKKTAPGRPG 344		
304 DPVNYQSGREGQSGIGYNLDHSEFNDFVFSFQONVRYTQLREKXY-----LVYTKNADAPA 357		

```
QY 345 CNTA-----DDKACAVGLGTEIKQKALAFDASYSRPRFLGNTANEFVIGADYNRPSTNEQ 400
D 358 TOTTLIRRPQK-----EENEISEFAIDNQLKATFATGVSNNHTVLGLDYKWL-----404
QY 401 GRTTLYARGGLALNEFRSPQVDLIANARKVGRVGSHTVATE-----NLDEFGIYKGS 453
D 405 -----TLEKMWLDNRNDYSFNWNPYV-----SVNDSMLTELSTNERNKLNQGVYLOD 456
QY 454 TPHPADGLSLIGGRLGHYKIESGE--GKTLHKASKTFTGYAGAVYDLNDNNSLYLSLS 511
D 457 QLE-WNQWNLILLSGRHWSRVDKQYAADTTTTERNDGKFTRAVRLLYAFDNGISPVVSYS 515
QY 512 QLYTQTNLDADG-KLLKPRQGNQFVGYKSGYMDRLNARVFRMKDKNAAPLNPN 570
D 516 TSFEPNLDGAPGTFAPKPTTGEQEVGVKFPKGSNTLLTVSLFDITQK-----ITSYN 571
QY 571 KTRXY-AALGKRVMEGVETEISGAVTPKQIIRHAGYSYLH-----SOIKTASNSRDD 620
D 572 SVTRNEQIGKVKSGVETEAHTQLTPEISLMAAYSITDAVTKESYASQVKNAPSS-----628
QY 621 GIELMPKHSANLWTTYYQ-----TPELTIGGGVNAWSGITSSAGMHAGGYATFDMAAYR 676
D 629 -----IPRHAASAWGSYSPHNGPLKGLTGTGVRYI-GSTTAITRRASRYRLYPVRRHGR 682
QY 677 P-----TPKLK--LQINADNIPNHHYARVGGANTFNIPGSERTWTANLRYSP 722
D 683 YELGSLASQLKGAQVQLNVLNLTDRHYVASCGR-DTACFYSGSGRTVATVSYSW 735

RESULT 12
FYUA_YEREN
ID FYUA_YEREN STANDARD; PRT; 673 AA.
AC P46360;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticin receptor precursor (IRPC) (IPR65).
GN FYUA.
OS Versinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51871 / WA-314 / Serotype O:8;
RX MEDLINE=95075311; PubMed=7984105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Versinia enterocolitica: a novel virulence
RT factor with dual function.";
RL Mol. Microbiol. 13:253-263 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1223-75-1, 8081, YE737, and 1209-79;
RA Rakin A., Heesemann J.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
CC SIDEROPHORE YERSINIACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29675; CAA82773.1; -
DR EMBL; Z35485; CAA84620.1; -
DR EMBL; Z35486; CAA84621.1; -
DR EMBL; Z35487; CAA84622.1; -
DR EMBL; Z35496; CAA84630.1; -
```

```
DR HSSP; P05825; IPEP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE NEG.
DR PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW Transport; TonB_box. 22
FT CHAIN 1 POTENTIAL.
FT SIGNAL 23 PESTICIN RECEPTOR.
FT SITE 30 TONB BOX.
FT SITE 37 TONB C-TERMINAL BOX.
FT SITE 657 TONB C-TERMINAL BOX.
FT VARIANT 362 N -> D (IN STRAINS 8081, YE737 AND 1209-79).
SQ SEQUENCE 673 AA; 73682 MW; 1C20E0352FAD4DCC CRC64;

Query Match 8.0%; Score 303; DB 1; Length 673;
Best Local Similarity 20.9%; Pred. No. 2.7e-13;
Matches 159; Conservative 98; Mismatches 298; Indels 204; Gaps 30;

QY 54 DYSSFAVTGKIPASUREIPQSIVSIITNQVKORNVDTFDQLARKTPGLRVLSNDDGR- 112
D 28 DESTLEVTASKQSSRSASANNVSTVSAPELSDAGVTASDKLPVLPLGLNI--ENSNGM 85
QY 113 --SSVYARGYEYSR--YN-----IDGLPAQMQSINGTLPNLFAPDRVENVMRGSGLPDS 162
D 86 LFTSISLRVGSQAQDFYNPAVTLVYDGVF---QLSTNTIQALTVDQSVELLRGPGQTLYG 142
QY 163 SGEAGGIVNLVYRKPTKAFQGHAAAGFGTHKQYKAEADVSGLSNSDGSVGRVNAQTVGA 222
D 143 KSAQGGIINIVTQOPDSTPRGYIEGGVSSRDSYRSKFNLSGPI-QDGLLYGSV----- 194
QY 223 SPRPAEKNNRHETFYAAADWD-INPDTVLGAGYLYQQRHLAPYNGLPADANKPLPSLPQH 281
D 195 -----TLRLQVDDGDMINPAT----- 210
QY 282 VFGADWNKFKWNSHDVFDLKHVFGNGG---YGVKVMYSDRSDADSNSYAFAGSKLGMK 337
D 211 ---GSD-----DL-----GGTRASIGNVKRLRAPDDQPWEMGFAASR----- 244
QY 338 TPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPRFLGNTANEFVIG- 391
D 245 -----ECTRATQDA-YVG-WNDIKGRKLSLSDGSPDPYMRRCCTDSQTLSCGYTTDDWVF 296
QY 392 NRPSTNEQRTTLYARGGLALNEFRSIPO-----VDLI-----AN 427
D 297 NLISAWQQOQHYSTRFPGSLIVN---MPQRWNQDVQELRAATLGDARTVDMVFLYRQN 352
QY 428 ARKVRG-----YSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESG-E 478
D 353 TREKLSNAYNMPYPLSSSTGYTTAETLAAYSDLTWHLTDRFDIGGGVRFSDHKSSTQYH 412
QY 479 GKTL-----HKASKTKFTGYAGAVYDLNDNNSLYLSLSQLYTPQ-----TNLDADGKL 526
D 413 GMLGNPFGDQKSNDDQVLGQLSAGYMLTDDWRVYTRIAQGYKPSGVNIPTAGLDAKP 472
QY 527 LKPRQGNQFVGYKSGYMDRLNARVFRMKDKNAAPLNPNKKTRYAALGKRVMEGV 586
D 473 FVAESKSYNELGTRYETADVTLOAATFYTHTKDMLQYS--GPVGMQTLNSA-GRADATGV 529
QY 587 ETEISGAVTPKQIIRHAGYSYLHSQI-----KTASNSRDDGIF-----LL 625
D 530 ELEAKRFPAGWSWDINGNIRSRSEFTNDSELYHGNRVFPVPRYAGSGSVNVIDTRYGAL 589
QY 626 MPKHSANLWTTYYQVTPELTIGGGVNAWSGITSSAGMHAGGYATFDMAAYRFTPKLQKI 685
D 590 MPRLAVNL-----VGPYFDGDN-----QLRQGTATLTDSSLSIGWQATERINISV 633
QY 686 NADNIPNHHY--YARVGGANTFNIPGSERTWTANLRYSP 722
D 634 HVDNLFDRRYRTYGYMGSSAVAQVNMGRVTGINTRIDF 672

RESULT 13
```

FYUA_YERPE STANDARD; PRT; 673 AA.
 AC P46359;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Pesticin receptor precursor (IRPC).
 GN FYUA OR PSN OR YPO1906.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OC NCBI_Taxid=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6/69;
 RA Rakin A., Heesemann J.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM6;
 RX MEDLINE=95204350; PubMed=7896707;
 RA Fetherston J.D., Lillard J.W., Perry R.D.;
 RT "Analysis of the pesticin receptor from Yersinia pestis: role in
 RT iron-deficient growth and possible regulation by its siderophore."
 RL J. Bacteriol. 177:1824-1833(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6/69;
 RA Buchrieser C., Rusnok C., Couve E., Frangoul L., Billault A.,
 RA Kunst F., Carniel E., Glaser P.;
 RT "DNA sequence of the 102 kDaes unstable region of Yersinia pestis."
 RT Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parthill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Bastam D., Bentley S.D., Brookes K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Fellwell T., Hamlin N., Holtroyd S., Jagers K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 CC - FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
 CC SIDEROPHORE YERSINIBACTIN.
 CC - SUBCELLULAR LOCATION: Outer membrane.
 CC - INDUCTION: BY IRON STARVATION. IRON REGULATION MEDIATED THROUGH
 CC THE PUR PROTEIN.
 CC - SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z55104; CAA84487.1; -;
 DR EMBL; U09530; AAA69906.1; -;
 DR EMBL; AL031866; CAA21395.1; -;
 DR EMBL; AL414150; CAC90722.1; -;
 DR HSRP; P05825; IFRP.
 DR InterPro; IPR000531; TONB_BOX.
 DR Pfam; PF00593; TONB_BOX; 1.
 DR PROSITE; PS00430; TONB DEPENDENT REC_1; FALSE NEG.
 DR PROSITE; PS01156; TONB DEPENDENT REC_2; FALSE NEG.
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
 KW Transport; TonB box; Complete proteome.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 673 PESTICIN RECEPTOR.

FT SITE 30 37 TONB BOX.
 FT SITE 657 673 TONB C-TERMINAL BOX.
 FT CONFLICT 484 484 G -> D (IN REF. 1).
 FT CONFLICT 514 514 R -> G (IN REF. 1).
 SQ SEQUENCE 673 AA; 73782 MW; 9C39E6010BCEB2C CRC64;
 Query Match 7.9%; Score 300; DB 1; Length 673;
 Best Local Similarity 20.9%; Pred. No. 4,4e-13;
 Matches 159; Conservative 97; Mismatches 299; Indels 204; Gaps 30;
 QY 54 DVSSFAVYVGTIKIPASLRIPQSVSIITNOQYKDRVDFPDQARTPLRLVLSNDGR 112
 DB DSTLLVVTASKSSRSASANNVSTVSAPELSDAVTSADKLPRVLPGLNI--ENSGNM 85
 QY 113 --SSVYARGYSE--YN-----IDGLPAQMSINGTLPLNFAFDRVEMRGPSPGLFDS 162
 DB LPTSLIRGVSSAQDPFYNAVTLVYDGV--QLSNTITQALTDVQSVELIRGPQSTLYG 142
 QY 163 SGEMGIYVLYVRKPTKAFQGAAGFTHKQYKADVSGSLSDSVGRVMAQTGVA 222
 DB 143 KSAQGIINIVTQQPDSTRGYIEGVSRSDSYRSKFNLSGPI-QDGLYGSV----- 194
 QY 223 SRPRAKNNRHETFYAADMW--INPDTVAGLYVQGRHLPNGLPADANNKLSLPQH 281
 DB 195 -----TLRQVDGDMINPAT----- 210
 QY 282 VFVQADNFKKNSHDFADLKHYPGNG--YGVKGRYSRDRADSNVAFASKLGMK 337
 DB 211 --GSD-----DL-----GSTRASIGNVKRLAPDDOPWEMGRFASR---- 244
 QY 338 TPAPRGCTADDKACAVGLGTEIKOKALAPASYSRPRLTANFEVIGA-----DY 391
 DB 245 -----ECTRAQDA-YVG-WNDIKGRKLSISGSDPDMRCCTDSQTLGKTTDWF 296
 QY 392 NRPRTNEGRRTTLVARGGLANERSIPQ-----VDLI-----AN 427
 DB 297 NISAWQOQHYSRTFPGSLIVN---MPQRNVQVQELRAATLGDARTVDVFGLYRON 352
 QY 428 ARGVYGV-----YSHVATENIDFQYKSTFHPADGSLIGGGLGHYKIESG-E 478
 DB 353 TREKLSAYDMTPMYLSTGTTTETLAYSDLTLMHLDTRPDIGGVAFSHDKSTQYN 412
 QY 479 GKT-----HKASKTKTYGAGAYVDLNNNSLYLSLQYTPQ-----INLADGKL 526
 DB 413 GSWLGNPFGDQKNDQVLLGQLSGVYMLTDWRVYTRAAQGVKPSGNVIVPAGDAP 472
 QY 527 LKPRQGNQPEVGYKSYMDRLNARVSPYRMDKRAAPLANPNKKTRYAALGKRVMEGV 586
 DB 473 FVAEKSYNELGTRYETADVTLQAATFYTHTRDMOLYS--GEVRMQTLISNA-GKADATGV 529
 QY 587 ETEISGAVTPKQOIHAGSVYLSQI-----KTASNRDDGIF-----LL 625
 DB 530 ELBAKRPAFGSMWDINGNVIRSEFTNDELYHGNRPVPRYGASSVNGVYIDTRYGL 589
 QY 626 MKHSANLWTTQVTPPELLTIGGVNAMSGITSSAGAHAGVATFPDAAAYRFTPKLQOI 685
 DB 590 MRLANL-----VGHYVDGN-----QLRQSTYATLDSLSLQWATERNNISV 633
 QY 686 NADNIFNRHY--YARVGANTNIPGSERTTANLRYSF 722
 DB 634 YVDNLFDRLRYRTGYWNGSSAVAQVMMGRTVGINTRIDF 672
 RESULT 14
 ID CIRA_ECOLI STANDARD; PRT; 663 AA.
 AC P17315;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Colicin I receptor precursor.
 GN CIRA OR CIR OR FEUA OR B2155.
 OS Escherichia coli.


```

Db          620 AMQVTKDVLTFRAGVTLNL 636

RESULT 15
PFEA_PSEAE ID PFEA_PSEAE STANDARD; PRT; 746 AA.
AC Q05098;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric enterobactin receptor precursor.
GN PFEA OR PA2688.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OK NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K407;
RX MEDLINE=93123148; PubMed=8419284;
RA Dean C.R., Poole K.;
RA "Cloning and characterization of the ferric enterobactin receptor gene (pfef) of Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:317-324(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Gahler R.L., Goltzy L., Tolentino E., Westbrock-Madhan S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latvig K., Lim R.M.,
RA Reizer J., Salier M.H., Hancock R.E.W., Loay S., Olsson I.T.,
RA "Complete genome sequence of Pseudomonas aeruginosa PAOI, an opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC ENTEROBACTIN.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY IRON AND ENTEROBACTIN.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch).
CC -----
DR DR EMBL; M98033; AAA25928.1; -.
DR EMBL; AB004697; AACG0676.1; -.
DR PIR; A40636; A40636.
DR HSPD; P05825; 1ERP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor; Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 746 FERRIC ENTEROBACTIN RECEPTOR.
FT SITE 39 44 TONB_BOX.
FT SITE 729 746 TONB_C-TERMINAL_BOX.
SO SEQUENCE 746 AA; 80967 MW; ADD4PB0CG22C3251 CMC64;

Query Match 6.8%; Score 256.5; DB 1; Length 746;
Best Local Similarity 21.3%; Pred.No. 4.9e-10;
Matches 175; Conservative 117; Mismatches 321; Indels 209; Gaps 40;

21 LSSSFAAQTDLEVVHKGGRSNALYTEKNGDYSSPAVVGTKIPSLKEIPDS--VS 78
| | : | | | | | : | | | | | : | | : | |

```

Dh	14	LSOCLL-----NAVAAAG-----	-DGSVTELEGEQTVATAQETKQAPGS	56
Qy	79	ITINQOVKDR-NVDTEPOLAKPTPGLRVLSND-----	-----DG-----	111
Dh	57	IITMEDIAKPPSNDLSQIIRITMEGVALTGNSSSGCGRNRRQIDIRMGENTLILVDGK	116	
Qy	112	-----RSSV-YA-REYEVSEVNIOSLPQMOSINGTLPNLPAFRVEMVRPSGLFDSGGE	165	
Dh	117	PVSSRNSVRYGMRGERSDRCGTNNVPADQ-----	-----VERIEYIRGPAAARYGNGA	164
Qy	166	MGGIVNLVRKRPYKAPFOGHA--AGFSTHK-----	-----QYAAEADVSG	203
Dh	165	AGGVNVIITQAGAEHTGNLSVSNFPGHKAEGASEMSRGLNGPLTENLSYVYGNIAK	224	
Qy	204	SLNSDGSVGRGVAQTVG--ASPPRAEKN-RHETFYAADWDINDPTVL--GAGYLQQ	258	
Dh	225	TDSDWDIMNAGHEENRFGKQAGTLPAAREGVNRKNDIGLISWRLTPRQTLFPAAGF-SRQ	283	
Qy	259	RHLAPYNGLPADANKPLPSLPQHYFVGADNNKFRMNSHDVPALDKHYFGNG-GYKAKGMR	317	
Dh	284	GNITTGDTQNTSNNNYKQW-----LGHETNRVRYETSYTHRGEMDFGSLAY----	LQ	334
Qy	318	Y-SRDADSNYAFPGSKLGMKTPAGRGCGCTADDKACAVGLAGTIEIKOKALAPASYSRPF	376	
Dh	335	YEKTRNRIKNEGLAGTGEIGFDP--NNAGFTTA-----	-----TLRDLTARGEVNLPL	380
Qy	377	RLGNTANEFYIADYNNFRSTNEQRTTLYARGLALNEFRSIPQVDLIANARKGYRGYS	436	
Dh	381	HLG-YEQTLTGLGSEWTRQKLDPPESNTQNTQEEG-----SIP-----	-----GLAGKN	422
Qy	437	HTVATENLDEGIFYGKSTFHPADGLSLGGGRIGHYKIES--GEQKTLHKASKTKFTGY	493	
Dh	423	RS-SSSSARIFSLPAEDNIEIMPETMLTPGLRWHDHDIYGDNNSPSLNLSHALTERTYTLK	481	
Qy	494	AGA-----VVDLNNNSLYLISQLYTPOTNL--DADGKLKPRQNGQFEVGYKGS	542	
Dh	482	AGIRAYKAPRYLQNLDPYLLISRGSGCTQSTISCYLRGNDG--LRAFTSVNKELGIE--	537	
Qy	543	YMDRLNARVSFYPMKDKN-AAAPLPD---NNKTRYA-----ALGKRVMEGYETE	589	
Dh	538	YSHDGLVAGLTPRYNDYKNNKIESGLSPVDHASGKGQYANAATUYQEMVNPYKAVVEGLE--	595	
Qy	590	ISGAVT-----PKWQHAGYSYIHSQIKTASNSRDDGILFMPKISANIATTYQVTPEL	643	
Dh	596	--GTTLTLPADGLKMSNNLT-----MLQSKKKEGTGDIVLTPRYATLNSMLMOQATDDL	647	
Qy	644	TIGGGVNAH-----SGITTSAGNHAGGYATFDMAAVRYRTPKXKLOINDNIF	691	
Dh	648	SLQATVWYWGKQPKKYDHGDRYTGANDQLSPYALAGIGTYRSLKNSLSTAGAVDNLF	707	
Qy	692	NRHYTARVG-----GANTFNIPGSERTWTANLRYSF	722	
Dh	708	DKRLP-RAGNAGVYVIGDAGAAATYNEPG--RFFYSLTNSF	746	
RESULT 16				
YE6A_HA6IN				
ID	YE6A_HA6IN	STANDARD;	PRT;	345 AA.
AC	086241;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical tomb-dependent receptor H11466.1.			
GN	H11466.1.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus.			
OX	NCBI_TaxID=727;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
FX	STRATIN-Rd / KM20 / ATCC 51907;			
FX	MEDLINE=9530630; PubMed=7542800;			
FA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			

```

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.,
RA Peterson J., Hickey E., Dodson R., Gwinn M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U32825; AAC23131.1; -.
DR TIGR; H11466.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Hypothetical protein; Receptor; TonB_box; Complete proteome.
FT SITE 325 345 TONB C-TERMINAL BOX.
SQ SEQUENCE 345 AA; 38744 MW; EAF83C55FCC9CB2A CRC64;

Query Match 5.9%; Score 221.5; DB 1; Length 345;
Best Local Similarity 24.1%; Pred. No. 4.2e-08;
Matches 82; Conservative 58; Mismatches 128; Indels 71; Gaps 16;

QY 435 YGHTVATE-----NLDFGIYKSTPHPADGLSLGGGRIGHY-----K 473
DB 25 YINPAIKNNENNAVQFNHLKTLGLYIGNTYFTDNFIITGLRYEYFDQVVGSRSLKN 84
QY 474 IESG-----EGTLHKASKTKTGTGAVYDNDNLSLSQLYPTQTN--LDADGKL 526
DB 85 IRSGVLAQDKGKLLYQ-----LGSYKFTPNIAFPFNHAEFRPNQNRRLINGE- 134
QY 527 LKPRGNOPEVYKGYMDRLNARVSFYRMKDKNAAPLNPNKKTRVYALGKRVMEGV 586
DB 135 LPAEQGSFETGLK--YENAYLNATVALENINKRNVAETVNVNG-TNELQIVGKQSRGI 191
QY 587 EPEISGAVTPKQIHHAGYSYLSQIKTASNSRD-----DGIFLLMPKHSANLWTTQ 638
DB 192 EFDLNGQLTDNLISIAANYTY--TKVKNLENHNKLVAGKQLSGV---PKHQASLFLAYN 245
QY 639 VTP-----ELTIGGVNAMSGITSSAGMHAGGY-----ATFDMAAYRFT---PKLKLQIN 686
DB 246 ICEFDFGNIRVGGGARYLGSWYANNNTYKAYLPOAIVDYDTFYADTKISGKVSFQLN 305
QY 687 ADNIENRHY-ARVGANTFNIP---GSERTWANLRYSF 722
DB 306 GKNLSNKKVYSPSTSGNASRTLIPVALGAREVILNTKIEF 345

RESULT 17
FATA_VIBAN STANDARD; PRT; 726 AA.
AC F11461; P19830;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric anguibactin receptor precursor (OM2).
GN FATA.

```

```

OS Vibrio anguillarum (Listonella anguillarum).
OG Plasmid pJMI.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.
OX NCBI_TaxID=55601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=775;
RX MEDLINE=88139336; PubMed=2830268;
RA Actis L.A., Tolmasy M.E., Farrell D.H., Crosa J.H.;
RT "Genetic and molecular characterization of essential components of
RT the Vibrio anguillarum plasmid-mediated iron-transport system.";
RL J. Biol. Chem. 263:2853-2860(1988).
RN [2]
RP SEQUENCE OF 637-726 FROM N.A.
RC STRAIN=775;
RX MEDLINE=90185247; PubMed=2311935;
RA Farrell D.H., Mikesell P., Actis L.A., Crosa J.H.;
RT "A regulatory gene, angR, of the iron uptake system of Vibrio
RT anguillarum: similarity with phage P22 cro and regulation by iron.";
RL Gene 86:45-51(1990).
CC -!- FUNCTION: ANGUIBACTIN RECEPTOR, ESSENTIAL COMPONENT OF THE IRON
CC UPTAKE SYSTEM OF V. ANGUILLARUM.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; J03529; AAA91581.1; -.
DR EMBL; M34504; AAA79859.1; -.
DR PIR; B29928; B29928.
DR PIR; PQ0051; PQ0051.
DR InterPro; IPR000531; TonB_boxC; 1.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB DEPENDENT REC 1; 1.
DR PROSITE; PS01156; TONB DEPENDENT REC 2; 1.
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW Transport; TonB_box; Plasmid.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 726 FERRIC ANGUIBACTIN RECEPTOR.
FT SITE 49 56 TONB BOX.
FT SITE 709 726 TONB C-TERMINAL BOX.
SQ SEQUENCE 726 AA; 78894 MW; 88B67A219395B154 CRC64;

Query Match 5.7%; Score 215.5; DB 1; Length 726;
Best Local Similarity 21.6%; Pred. No. 3.1e-07;
Matches 163; Conservative 115; Mismatches 338; Indels 137; Gaps 37;

QY 9 RINNTAATVLAALS---SSVFAAQTADLETVHIKQSYNAIVTEKNGDYSSFAVTVGCK 65
DB 22 RIALVALQISASALPISITHAEQAD-ESITVYQGAN-EAYAAGKISKASSIGH-LGDK 78
QY 66 IPASLREIPQSVSITITQQVKDRNV-DTFDQARKTKPLRLVLSNDG--RSSVYARG-- 120
DB 79 ---DFLDTFPNAIGYTDKHQIQDHAQDISVISASDSPV-FTSGETGLNKESPKIRGSS 134
QY 121 EYSEYNIDGL-----PAQMQSINGTLPLNLPFADRVENVRGPSGLFDS---SGEMGGIVNLV 173
DB 135 DIGDVMFNGLYGIAPIYRSS-----PEM--YQRIDVLKGPASLLNGMPNGSVGGINLV 187
QY 174 KRPTKA-----FQGHAAAG--FGTHKQYKAED---VSGSLNSDGSVRGR 214
DB 188 TKRAQEAIPITFTGYMSDSQFGGHIDIGRRFGENEGVGFVFRDGDASVQGSRKA 247
QY 215 VMAQTGVGASPRPAEKNRHHETFYAAADWDINPDVTGLGAGLYQQRHLAPYNGLPADANNK 274
DB 248 QLAS-----LSLDWRNDIALIEADLYFSTERVVDGPNRGLSTASGVD 288

```

```

QY 275 LPSLP-QHYFVGADW--NFKRNKSHVFPADLKHFGNGCYGKGYKRYSDRDADSNATPAG 331
DB 289 VSPSPSDTLTSPMAYNSEDEKMMIRAEILDSNSVTAYGAVGA--SRTDFSNVPORY 346
QY 332 SKLGKMTAPAGPCPNADKACAVIGT--BIKKALAFDASYSRPRLTANEPIYAGD 390
DB 347 KIID-----DSGTLEVSLSGVALKESKRTSGEIGIRSSPDGTGHEHYLVNST 393
QY 391 YNFRSTNEGRTTLTYARAGGLALNEFRSI--POVDLIANARKGVRGSHVATENIDEPG 448
DB 394 Y--FRDDKNDSPFGNNPGSMNINYPWGPEDSTYDN-----YELPVDSTQISFG 444
QY 449 IYKSTFHPADGSLIGG-----RIGHYKIESGEGKTLH-----KASKTFYGYAGA 496
DB 445 V-----ADTLISANGKQVTLGLRHOSIDYEGVTWNGNAFPPTKLSESTYPAIVA 496
QY 497 VYDLNNNSLYLSLSQLYTPQNTLAD-----GKLLKPROQNGPEVGYKSGYMDRLNARY 552
DB 497 LYKVSUSVSLYGYTGLTNGKTAGSGAANVGAPEPQKYTEAGIKLD--MNDPAHT-F 554
QY 553 SFYRMKDKNAAAPLNPNNKTRRYALGKRVMEGVEITEISGAVTPKQIHAGSYLHSQIK 612
DB 555 SLEIETKPKGYODPDN-----IYSPGEQRNRIEMGFYGTVLDTLTGIAVTDAEIT 610
QY 613 TASNRSDDG-IFLLMPKHSANL--WTTVQVTELTIGGCVNAMSQ--ITSSAGMAGY 666
DB 611 KATDVTEGKQATKLPDLQKALAEW--NLPMVQLTLIGQANMSQYIDAQNTQSLSAQ 669
QY 667 ATPDAAAYRFT---PKLKIQINADNIFNRHY 696
DB 670 TIFDLGARYNSTIANOSVIRLAVNNVTBAYW 702

RESULT 18
IRGA_VIRCH STANDARD; PRT; 652 AA.
ID IRGA_VIRCH STANDARD; PRT; 652 AA.
AC P27772; Q9KUP0;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated outer membrane virulence protein precursor.
GN IRGA OR VC0475.
OC Virbio cholerae.
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=classical Ogawa 395 / ATCC 39541 / Serotype O1;
RA MEDLINE=93023868; PubMed=1406279;
RA Goldberg M.B., Boyko S.A., Butterton J.R., Stroebner J.A.,
RA Payne S.M., Calderwood S.B.;
RT "Characterization of a Vibrio cholerae virulence factor homologous to
RT the family of TonB-dependent proteins.";
RL Mol. Microbiol. 6:2407-2418(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ustebach T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Frazer C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
RN [3]
RP SEQUENCE OF 1-151 FROM N.A.
RC STRAIN=classical Ogawa 395 / ATCC 39541 / Serotype O1;
RA MEDLINE=91072235; PubMed=2174861;
RA Goldberg M.B., Boyko S.A., Calderwood S.B.;

```

```

RT "Transcriptional regulation by iron of a Vibrio cholerae virulence
RT gene and homology of the gene to the Escherichia coli fur system.";
RL J. Bacteriol. 172:6863-6870(1990).
CC -1- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING
CC FERRIC VIBRIOBACTIN. AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC VIBRIOBACTIN TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: REGULATION OF THE IRGA EXPRESSION IS NEGATIVELY
CC REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, U72152; AAC4766.1; -.
CC EMBL, AE004134; AAF93648.1; -.
CC PIR, A37834; A37834.
CC PIR, S25265; S25265.
CC HSRP, P03825; IFEF.
CC TIGR, VC0475; -.
CC InterPro, IPR000531; TonB boxC.
CC Pfam, PF00593; TonB boxC.1.
CC PROSITE, PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE, PS0156; TONB_DEPENDENT_REC_2; 1.
CC Virulence, Outer membrane, Iron transport; Transport; TonB box;
CC Signal; Receptor; Complete proteome.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 652 IRON-REGULATED OUTER MEMBRANE VIRULENCE
CC PROTEIN.
CC SITE 33 40 TONB BOX.
CC SITE 635 652 TONB C-TERMINAL BOX.
CC CONFLICT 294 294 D -> G (IN REF. 1).
CC CONFLICT 448 448 K -> Q (IN REF. 1).
CC CONFLICT 502 502 A -> T (IN REF. 1).
CC SEQUENCE 652 AA, 71669 MW, A95F82FECC72EC93 CRC64;

Query Match 5.7%; Score 214; DB 1; Length 652;
Best Local Similarity 19.6%; Pred. No. 3.4e-07;
Matches 146; Conservative 98; Mismatches 305; Indels 196; Gaps 31;

QY 57 SPAYVTGKIPAS-----LREIPQSVGIIITNOQYKDRVDF 93
DB 10 SLSTVLGLMFSASAPFAQADTKTDETWVTAAGYAQVIONAPASISVISREDELSRYRDV 69
QY 94 DQIARKTPLRLVLSNDDGRS-SVYARGVEYSEYNIDGLPAQMOI-----NGTLP 142
DB 70 TDALKVPEVTVYGGDDTTDISIRGMSNYTLIVDGKQTSRQRPNSDPRIEGQWLP 129
QY 143 NLEAFDREVMRPGSLFDSGEMGIVLVKRRPTKAFQGHAAAGFTHKQYKADAVS 202
DB 130 PLQAIIRIEVIRGPMSTLVGSDAIGVINITRKQOQSGNVQJSTVVQENRASDEQS 189
QY 203 GSINSGSVRGRMAQTVASRPAKNNR-----ETFAAAMDNDNP--TVLGAG 253
DB 190 ANEFTVGPLSDALSLQVYQTTQDEDEIEHGYSKLSLTSKLVYQINPHQQLLEAG 249
QY 254 YLYQQR-----HLAPNGLPADANNKLPSPQHVFG--ADWNKFKMNSHVPADLK 303
DB 250 VSAQDEBNVNGSAGSSGCGRTCSNTDQYRR--NHAIVSHQGDW-----QDV----- 295
QY 304 HFFGNGGYGVKMGVSDRDADS-----NYAFAGSKLGMKTPAGRGPCNTADKACAVGL 357
DB 296 -----GQSDTYLYQYEBENTNKSREMSIDNTVF--KSTLVAFIGE--MLSF 336
QY 358 GTFIKKALAFDASYSRPRLTANEPIYAGDYNRFRSTNQGRITLYARGLALNEFR 417
DB 337 GVGKKEHSLE-----DKTSNKI-----SSRTHLSNTQW-----AGFIDEWA 373

```

QY 418 SIPOVDLIANARKGVGYSHYVATENLDFEGLYKSTFHP-----ADGLSLGGRLG 470
 Db 374 LAEOPLRTGGR-----LDHKNYG-SHFSRPRYGVWNLDPLTWVRGG--- 415
 QY 471 HVKISBG-EGKTLHKASKT--KFTGVAGAVY--DLNDNNSLYLSLSOLYTPQTNLDADG 524
 Db 416 ---VSTGFRAPOREVTPWQVSG-GGNIYGNPDLKPETSKINKLSLMYSTGSLAAS- 470
 QY 525 KLLKPRQGNQFEVGYKSGYMDRLNARVYFRMKDKNAAPLN-----PNN---KKTRYA 576
 Db 471 -----LTAFHNDKDKITRVACPANICTAGBNQWGAAPTYYR 507
 QY 577 ALGKRVMEGVETEISGAVTPKKQIHAGYSLHSQIKTASNSRDDGIFLLMPKHSANLWTT 636
 Db 508 NIDEAETVGAETLSLPITESVELSSSYTYTHSEQKSGNFAGR-LLQLPKHFLNANLS 565
 QY 637 YQVTPTELTIGGVN-----ANSGITSSAGMHAGGVATFDAMAARFYPKCLKLOINADNI 690
 Db 566 WQTDRLNSWANLVYRKEMQEGGASNDDFIAPSYTFIDTGVYALTDTATIKAAVYNL 625
 QY 691 FNRHY-YARVGGANTFNIPGSRTW 714
 Db 626 PQEYNYAEVG-----YVEDGRYV 645

RESULT 19
 YNCD_ECOLI
 ID_YNCD_ECOLI STANDARD; PRT; 700 AA.
 AC P76115;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable TonB-dependent receptor yncd precursor.
 GN YNCD OR B1451.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000242; AAC74533.1; -;
 DR EcoGene; EG13774; YncD.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
 KW Hypothetical protein; Outer membrane; Receptor; Signal; TonB box;
 KW Complete proteome.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 700 PROBABLE TONB-DEPENDENT RECEPTOR YNCD.
 FT SITE 680 700 TONB C-TERMINAL BOX.
 SQ SEQUENCE 700 AA; 77260 MW; 7B3B96C6ABA48FE4 CRC64;

Query Match 5.78; Score 214; DB 1; Length 700;
 Best Local Similarity 20.0%; Pred. No. 3.8e-07;
 Matches 156; Conservative 114; Mismatches 317; Indels 192; Gaps 31;
 QY 5 MSVFRINWTAATVLAALSSSVFAAQTADLETVHIKGRSNAIVTEKNGDYSSFAVTVGT 64
 Db 1 MKIFSVRQTVLPALLVLSPVVFAA-----DEQTVIVSAAP 35
 QY 65 KIPASLREIPOSVIITNQVVK--DRNVDTFDQLARKTPGLRVLNSDD----- 110
 Db 36 QWSEL-DTPAAVSVVDGEEMRLATPRINLSESLT-GVPGQLQVQRNQRYAQDLQLSIRGP 93
 QY 111 GRSSVYARGYSEYNVDGLPAQMSINGTLPN--LFAFDREVVRGP-SGLPFSSSGMG 167
 Db 94 GSRSTY--GIRGIRLYVDGIPATMPDGGQTSNDLSVQNVVEVLGRFPFSALYNAS--G 149
 QY 168 GIVNLVRKPTKAFQGHAAAGFGTHKQYKAEADVSGSLNSDGSVGRVMAQTVGASPRPA 227
 Db 150 GVMNVTTQTGGQPTIEASSYSGFSGWRVYGLKATGA-TGDTQTPGDV----- 196
 QY 228 EKNRHETFYAAADWDINPDTVLGAGYLYQ---QRHLAPYN-GLPADANNKPLSPQHVF 283
 Db 197 -----DYTVSTTRFTTHGYRDHSGAQKLANAKLGVRIDEASKLSLI----- 238
 QY 284 VGADWKEFKMNSHDVFPADLKHYPGNGGYKVMRYSDRSDANSYAFAGSKLGMKTPAG-- 341
 Db 239 -----FNSVDIKAD-----DPGGLTKAEWKANPQAPRAEQYDTRKTKIKTOAGLR 284
 QY 342 -RPGCNTADDKACAVGLG-TEIKOKALAFDASYSRPFRLGN--TANEFVIGADYNNRPRST 397
 Db 285 YERSLSRDDMSVMYAGERETTYQSIHPAQLNPSHAGGVITLQRHYQID-SRWTHR 343
 QY 398 NEOGRTLYARGGLALNEFRSIPQVDLIANARKGVGYSHYVATENLDFEGLY 451
 Db 344 GELGVPTFTTGT---LN-----YENMSENKRGYNNPRLNSGMPYQKGLERDE 390
 QY 452 -----KSTFHPADGLSLGGRLG-----HYKIESGEGKTLHKASKTKFTG 492
 Db 391 RNLMMNIDPYLQTMQLSEKLSLDAGVRYSSVWFDSDHY-VTFPGNGDSDGASVHKWLP 449
 QY 493 YAGAVYDLNDNNSLYLSLSOLY-TPQTN---LDADGKL-----LKPRQGNQFEVGYKSGY 543
 Db 450 AGSKYAMTDANNIYLAAGRGFTPTNELSYRADGGSGMNLGLKPSNTDITIEGSKTRI 509
 QY 544 MDDLRLNARVSPYRMDKNAAPLNPNKTKRYAALGRVMEGVETEISGAVTPKWQIHAG 603
 Db 510 GDGLLS--LALFQDITDDEIIVDSSSGRTTYKNAKTRRQGAELAWDQRFAGFRYNAS 567
 QY 604 YSYLHSQIKTASNSRDDGIFLLMPKHSANLWTTVQVTELTIGGVNAMSITSSAGMHA 663
 Db 568 WTWLDATYRSNVCNEQDCNGRMFGIARNMGF-----ASIGYVPEDGMYA 612
 QY 664 GGYATF--DAMAARFPTPK-----LKLQINADNIENRHYARV 699
 Db 613 GTEARYMGDIMADDENTAKAPSYTLVGLFTGYKYNHNLTVDLFRVDNLFDEYVGSV 671

RESULT 20
 BTUB_SALTY
 ID_BTUB_SALTY STANDARD; PRT; 614 AA.
 AC P37409;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Vitamin B12 receptor precursor.
 GN BTUB OR STM4130.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=L72;

RX MEDLINE-93079446; PubMed1448622;
 RA Wei B.Y., Bradbeer C., Kadner R.J.;
 RT Cyserved structural and regulatory regions in the Salmonella
 RT typhimurium trbB gene for the outer membrane vitamin B12 transport
 RT protein.";
 RL Res. Microbiol. 143:459-466(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE-21534948; PubMed1167609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -!- FUNCTION: COBAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
 CC BACTERIOPHAGE BP23 AND IS NECESSARY FOR THE UPTAKE OF E. COLICINS.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M89481; AAA27031.1; -
 DR EMBL; AE008893; AAL22968.1; -
 DR StyGene; SG10030; trbB.
 DR InterPro; IPR000531; TonB boxC.
 DR Pfam; PF00593; TonB boxC_1.
 DR PROSITE; PS00430; TONB DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB DEPENDENT_REC_2; 1.
 KW Outer membrane; Cobalt transport; Transport; TonB box; Signal;
 KW Receptor; Complete proteome.
 FT SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 1 614 VITAMIN B12 RECEPTOR.
 FT SITE 26 33 TONB BOX.
 FT SITE 597 614 TONB C-TERMINAL BOX.
 FT CONFLICT 81 81 G -> R (IN REF. 1).
 FT CONFLICT 134 134 R -> P (IN REF. 1).
 FT CONFLICT 183 183 N -> I (IN REF. 1).
 FT CONFLICT 280 280 R -> S (IN REF. 1).
 SQ SEQUENCE 614 AA; 68525 MW; 9F51F601A615FD62 CRC64;
 Query Match 5.4%; Score 205; DB 1; Length 614;
 Best Local Similarity 21.1%; Pred. No. 1.3e-06; Indels 206; Gaps 39;
 Matches 161; Conservative 115; Mismatches 280;

QY 285 GADWNNKFNNSHDV--FADLKHVFGNGYGVKVMRYSDRSDASNAFAAGSKLGMKTPAGR 342
 DB 273 GHHFNGERLQSQVSSYSHSKDYNDPHYG-----RX-DTSA----- 308
 QY 343 PGCNTADDDKACAVAGLGEIKOKALAFDASYSRPFRLGNTANEFGADYNNRFRST----- 397
 DB 309 -----TLD-----ENKO-----YNNQWTSVVGH--GNVGAGVDNQKOTTPGTGY 348
 QY 398 ----NEQGRITLYARGGLALNFRSIPQVDLJANARKGVRGSHVATENLDFEFGYGS 453
 DB 349 VPEGVDQKNTGYVLTGLQGLDF-----TLEMAAR-----SDNSQFGRHG-- 389
 QY 454 TFPADGLSLGGGRIGHYKISBEGKTLHKASKTFGTGAGAVYDLNDSLSLISQL 513
 DB 390 TWGTSAGWPFIS-----YRPLASYG-TSYKAP-----NLGGL 421
 QY 514 YTPQTNLADGKLLKPRQGNQPEVGYKSYMDRLNARVSFPRMDKNAAPLNPNKKT 573
 DB 422 YGYGPN-----LNPEKSKWEGAFEG--LTAGVSWRISGYR-NDINMIDYDHLQK- 472
 QY 574 RYAAAGKRWMEGVETEI---SGAVTPKQIH--AGVYLHSQIKTASNSRDDGIFILMPKH 629
 DB 473 -YNEBKARIKGLEATNFTGPLET-----HTVSYDYV-----DARNALTD---TPLPR 518
 QY 630 SANL-----WTTYQVTPELT---IGGVNAMSGLITSSAGMHAGYATPDMAAAYRFTPK 680
 DB 519 SKQMAKYQLDMDVDPDWMGTQYIGSRSDSYSAAPRYTVMKGVSIMDLTYAVYVTS 578
 QY 681 LKQLINADNIPNRHYYARVGANTENIPGSERTWTANLRYSP 722
 DB 579 LTVRGKIANLFDKDYETVYG---YQTAGREYTLGSS--YTF 614
 RESULT 21
 FECA_ECOLI STANDARD; PRT; 774 AA.
 ID AC P13036;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Itron(III) diclitate transport protein fecA precursor.
 GN FECA OR B4291.
 OS Escherichia coli.
 OC Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OK NCBI_TaxId=562;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-40.
 RC STRAIN=B;
 RX MEDLINE=88227855; PubMed=2836368;
 RA Presler U., Staudenmaier H., Zimmermann L., Braun V.;
 RT "Genetics of the iron diclitate transport system of Escherichia
 RT coli.";
 RL J. Bacteriol. 170:2716-2724(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=95334362; PubMed=7610040;
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
 RA Blattner F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes.";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [3]
 RP SEQUENCE OF 1-428 FROM N.A.
 RX MEDLINE=91072220; PubMed=2254251;
 RA van Hove B., Staudenmaier H., Braun V.;
 RT "Novel two-component transmembrane transcription control: regulation
 RT of iron diclitate transport in Escherichia coli K-12.";
 RL J. Bacteriol. 172:6749-6758(1990).
 RN [4]
 RP SEQUENCE OF 731-774 FROM N.A.

RC STRAIN=K12;
RX MEDLINE=89213950; PubMed=2651410;
RA Staudenmaier H., van Hove B., Varaghi Z., Braun V.;
RT "Nucleotide sequences of the fecBDE genes and locations of the
RT proteins suggest a periplasmic-binding-protein-dependent transport
RT mechanism for iron(III) dicitrate in *Escherichia coli*.";
RL J. Bacteriol. 171:2626-2633(1989).
CC -!- FUNCTION: FECA is the OUTER MEMBRANE RECEPTOR PROTEIN IN THE
CC IRON(III) DICITRATE TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
CC AND INDUCTION BY THE IRON LEVEL MEDIATED BY THE FUR PROTEIN
CC DICITRATE COMPLEX MUST ENTER THE PERIPLASM WHERE IT BINDS TO A
CC TRANSMEMBRANE PROTEIN, WHICH REGULATES DIRECTLY OR VIA A FURTHER
CC INDUCTOR, TRANSCRIPTION OF THE FEC GENES.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20981; AAA23760.1; -;
DR EMBL; U14003; AAA97187.1; -;
DR EMBL; AE000499; RAC7247.1; -;
DR EMBL; M63115; AAA23768.1; -;
DR EMBL; M26397; AAA23761.1; -;
DR PIR; JY0022; QRECEA.
DR EcoGene; EG10286; fecA.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 33
FT CHAIN 34 774
FT SITE 56 63
FT SITE 757 774
FT CONFLICT 16 16
FT CONFLICT 190 190
FT CONFLICT 300 300
FT CONFLICT 357 357
FT CONFLICT 444 444
FT CONFLICT 749 749
SQ SEQUENCE 774 AA; 85321 MW; 16B5B510276C3B09 CRC64;

Query Match 5.4%; Score 205; DB 1; Length 774;
Best Local Similarity 20.1%; Pred. No. 1.8e-06;
Matches 170; Conservative 108; Mismatches 316; Indels 250; Gaps 39;

QY 18 LAALSSVFAAQ-----TADLETVHIKQORSYNAIVTEKNGDY- 55
DB 24 LAGLSFAFAAQNTAPGSLDKALNQYAAHSGFTLSVDASLTRGQS-NGL-----HGDYD 78
QY 56 -----SSPAVTV-----GKTPASLREIPQSVS 78
DB 79 VESGLQLLDGSLQVKPLGNNSWTLEPAPAPKEDALTGVGDWLDARENDVFEHAGARD 138
QY 79 IITNQVKDRNDVTFDQARKTGLRLVSLNDDGRSSVYARGVESEYN-----ID 128
DB 139 VTRREDFAKTATWREVLNRIIPGVSAPENNNGTGSDDLAWNFGLRLNRLASRSTVLMD 198
QY 129 GLPA-----QMSINGTLNLFADFVEVMRPGSLFDSSGEMGGIVNLVRKRPTKA 180
DB 199 GIPVPFAPYQQLSLAPVSLGNM---DAIDVVRGGAVRYGVSQVGVVNFV---TRA 251
QY 181 FQGHAAAGFTGTHQYKAEADVSGSLNSDGSVGRVVAQTVGASPRPAEKNNRHET----- 235

Db 252 IP-----QDFGIEAGVEGQLSPSTSS-----QNNPKETHNLMV 283
QY 236 -----FYA---AADWDIN-----PDTVLGAGLYQQRH-----LAPYNG---L 267
DB 284 GGTADNGFGTALLVSGTRGSDMREHSATRIDDLMLKSKYAPDEVHTFNSLLOYDGEADM 343
QY 268 PA-----DANNKLPRLPQHVFGADWNKPKFMS--HDVFADLKHVFGNGGYGKVGWRY 318
DB 344 PGLSLRADYDADRQWQSTRPYDRF-----WGRRLKASLGLGQFQDSQHKFNIGQFYQTILR- 398
QY 319 SDRADSNYAFAGSKLGMKTPAGPGCGCTADDKACAVGLGTEIKQKALAPDASVSRPRL 378
DB 399 -----SGYLEQGRKITL-----SPRNYWVRGIEPRYSQIFMI 430
QY 379 GNTANEFVIGADYNFRSTNEQGRITLYARGGLALNEFRSIPQVDLIANARKVGRGYSHT 438
DB 431 GPSAHEVGVGRYLN-ESTHEMRYTATSSQLPSG---SSP---YDRDTRSGTEAHAW- 482
QY 439 VATENLDEFYIGKSTPHPADGLSLIGGRLGHVYKISGEK-----TLHKASKTKFTGYA 494
DB 483 ----YLDCKIDIGNWTTP-----GMRPEH--IESYQNNAITGTHEEVSYNAPLAL 528
QY 495 GAVYDLNDNNSLYL-----SLSQLYTPQTNLDADCKLLKPRQGNQFVYKSGSYMDRLNA 550
DB 529 NVLYHLTDSMNLVYANTGESFGTVQYQSLGKAVQSGNVEPEKARTWELGTR--YDDGALTA 586
QY 551 RVSPYRMKDKNAAAPLNPNNKKTRYAALGKRVMEGVETEIS---GAVTP---KWQIHAGY 604
DB 587 EMGLFLINPNN---QYDSNQNTNDVTARGKTRHTGLETOARYDLGTLTPTLDNVSIYASY 643
QY 605 SYLHSQIKTASNSRDDGIFLLMPKHSANL-----WTTYQVTPTELTIGGVNAMSGIT 656
DB 644 AYVNAEIREKGDITYGN-LVPPSPKHKGTGLVDYKPGNW-TFNLSDFGSSQFADNANTVK 701
QY 657 SSAGMHAG--GYATFDMAAYRTPK---LKLQINADNIENRHYRYARVGANTFNI-PG 709
DB 702 ESADGSGTRIFGFMWGARVAIDFGPQWADNLAFGVKNIPFDQDYFIRSYDDNNKGIYAG 761
QY 710 SERT 713
DB 762 QPRT 765

RESULT 22
HEMR YEREN
ID -HEMR YEREN STANDARD; PRT; 687 AA.
AC P31499;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hemrin receptor precursor.
GN HEMR.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA MEDLINE=93049186; PubMed=1425573;
RA Stojiljkovic I., Hantke K.;
RT "Hemin uptake system of Yersinia enterocolitica: similarities with
RT other TonB-dependent systems in Gram-negative bacteria.";
RL EMBO J. 11:4359-4367(1992).
[2]
RP REVISIONS.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA Stojiljkovic I.;
RL submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.

```

CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X68147; CAA48250.1; -
DR PIR, S28042; S28042.
DR InterPro, IPR000531; TonB_boxC.
DR Pfam, PF00593; TonB_boxC; 1.
DR PROSITE, PS00430; TONB DEPENDENT REC 1; 1.
DR PROSITE, PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal;
KW Receptor.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 687 HEMIN RECEPTOR.
FT SITE 44 51 TONB BOX.
FT SITE 670 687 TONB C-TERMINAL BOX.
SQ SEQUENCE 687 AA, 75226 MW, 69092EP97F5DC08A CRC64;

```

```

Query Match 5.3%; Score 200; DB 1; Length 687;
Best Local Similarity 20.3%; Pred. No. 3.3e-06;
Matches 158; Conservative 96; Mismatches 322; Indels 204; Gaps 32;

```

```

QY 8 PRINMTAATVLAALSSVPAQAQDADETHVHIKQGRSYNAIVTEKNDGYSSFAVTVGTRK 67
DB 9 FPMSPSLIACTSLAVQAADATSSSTQ-NSSKRIADTVMTATGERSSSF----- 58
QY 68 ASLRREIPQSVSIITNQVDRNVDTFDQARKTPGRLVLSNDGR--SSVARGEYSE 124
DB 59 ----EAPMNVTVBEATPTPTSETATSDMLKRIPLGLTVGS--GRVNGQDVTLRIGKQG 112
QY 125 Y--NIDGLPAQOMQS--INGTLPNLFAFDRVVRKSPGSGFDSSGEGGIV-----NL 172
DB 113 VTLTVGDIHQGTDTGHLNSTFLDPAVKRVEIVRSPALLYSGGALGGVSYETVDAADL 172
QY 173 VKRPKPAQGHAAAFGRH-----KQYKADVSGSLN---SGSVARGVMAQTGA 222
DB 173 LPPGQNSGKRVYSAATGSHSPGLGASAFKRTDVGIIISFGTRDIGNIR-----GSDGF 227
QY 223 SPRPAKNNRHETFYAADMDINPDTVLGAGVLYQGRHLAPYNGLPADANNKLPSPQHV 282
DB 228 N--AENDETISVYLAKGTWRIDQIQLSALNRY-----YNSALEPKNPOTSAASST 277
QY 283 FVGADNNKFRKNSHDVFADLKHYFGNGYGVKVMARYSDRDADSNVAFASGKLGK----- 337
DB 278 NMTDRSTLIQRDAQ-----LKNIKRPLDEWMLNATQVYVSEIN 318
QY 338 -TPAGPPGNCNTADDKACAVGLTEIKOKALAPDASYSRPRLGNT-----ANER 385
DB 319 AAPQGTPT-----EERGKQTTKGKGLKRTLTFTDSFASHLLTGTBAVYKQGTTPSGATES 373
QY 386 VIGADVNRFRSTNEGRTTLTYAGGLALNEFRSIPQVDLIANR-KGVYGSITVATENL 444
DB 374 FPQADL-KRFGSGLQDEITL-----RDLP-VSIIAGTRYNNYRSSSEGVADVDA 420
QY 445 DEFGIYKSTFHPADGSLILGGRLGHYKLESGEKTLHKASXTKFTGYA-----GA 496
DB 421 DKSSSRGAVSVPTDWMLEFG-----SYAQAFRAPTMBE 454
QY 497 VYDLADNNLSLYLSQLYTPQTNLDADGKLKPRQGNQFEVYKGSYMD-----DRINAR 551
DB 455 MY-----NDSKHFEMNIMGWTLITNYWPNPILKEBTNETQEGYGLFNDLMAEDDLQFK 510
QY 552 VSPYRKMDKNAAPLNPNNKRTFYALGKRVMEG-----VEIEISGATTPPKV 558
DB 511 ASIF--DTNA-----KDYISTGVTMDFGFGPGGLYCKNCSTYSTINIDRAKIWM 557

```

```

QY 599 QIHAGSYHSQIKTASNRDQIFLLMPKHSANLW-----TTYQVPELTI GGAVNM 652
DB 558 DATMTYQTMFNLGLAYN-KTRG-----XQONTNEWLDTPNIDPTVSTDDVPANGFAV 611
QY 653 SGI-----TSSAGMHAGVATFDMAAYRFPKTK---LQINADNIPNNHYARVG 700
DB 612 GWIGTRADRSRVSSSGTPQAGVGNDFVSYKGEQFKGMTTTVVLGNAPDKYIGPQG 671

```

RESULT 23

```

ID TUTA_ECOLI STANDARD; PRT; 732 AA.
AC P14542;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ferric aerobactin receptor precursor (Cloacin receptor).
GN TUTA.
OS Escherichia coli.
OG Plasmid IncF ColV3-K30.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F344;
RA Krome W.J.A., Stegehuis F., Koningsstein G., van Doorn C.,
RA Roosendaal B., de Graaf F.K., Oudega B.;
RT "Characterization of the pColV-K30 encoded cloacin Df13/aerobactin
RT outer membrane receptor protein of Escherichia coli; isolation and
RT purification of the protein and analysis of its nucleotide sequence
RT and primary structure."
RL FEMS Microbiol. Lett. 26:153-161(1985).
RN [2]
RP REVISIONS.
RA Oudega B.;
RT Submitted (FEB-1992) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR CLOACIN DF13/AEROBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X05874; CAA29297.1; -.
DR EMBL, X05874; CAA29298.1; ALT_INIT.
DR PIR, S01042; S01042.
DR InterPro, IPR000531; TonB_boxC.
DR Pfam, PF00593; TonB_boxC; 1.
DR PROSITE, PS00430; TONB DEPENDENT REC 1; 1.
DR PROSITE, PS01156; TONB DEPENDENT REC 2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Plasmid;
KW Receptor.
FT SIGNAL 1 25
FT CHAIN 26 732 FERRIC AEROBACTIN RECEPTOR.
FT SITE 31 38 TONB BOX.
FT SITE 715 732 TONB C-TERMINAL BOX.
SQ SEQUENCE 732 AA, 81014 MW, 0C23879C0B27AE2B CRC64;

```

```

Query Match 5.2%; Score 198; DB 1; Length 732;
Best Local Similarity 19.5%; Pred. No. 5e-06;
Matches 174; Conservative 116; Mismatches 256; Indels 346; Gaps 48;

```

```

QY 6 SVRRIMTAATVLAALSSVPAQAQDADETHVHIKQGRSYNAIVTEKNDGYSSFAVTVGTRK 65
DB 8 TLWALNPILLITMADPA-----VAQQTDDTFVVASNRSKRYA----- 45
QY 66 IPASLRREIPQSVSIITNQVDRNVDTFDQARKTPGRLVLSNDGRSSVYARGEY 122

```

```
Db 46 -----EMAQTWVIEAELQIOGGKELKDALAQIPGLDVSRSTNYGMNVGRPL 99
Qy 123 SEYNDIGL-----PAQMQINGTLENLFAFDEVRVNRGSGIFDSSGEMGGIVNLVR 174
Db 100 VVL-VGVRLNSSRDSRLQSDIDP-----FNWHIEVIFGATSLY-GGGSTGGLINIVT 152
Qy 175 KR-----PTKAFQGHAAAGTGTHK-----
Db 153 KKGQPTWMEFEAGTKSGFSSSKDHERIAGAVSGGNEHISGRLSVAYQKFGGFDGNGD 212
Qy 194 -----QYKAEADV--SGSLNSDGSVRGVMAQTVGASPRPAEKNNHEHFFYAAA 240
Db 213 AFLDNTQTGLQYSDRLDIMGTTINIDESRLQITQ-----YKQS 255
Qy 241 ---DWDINPDTVLGAGLYQOORHLAPY--NGLPADANNKLPSLPQHVFGADVANKFKMNS 295
Db 256 GDDVGLN-----LGKGFSAIRGTSTFPVSNGLNSD---RIFGTGHL----- 295
Qy 296 HDVFADLKHFGNGGKVGKMYRSDRSDSNYAFAGSKLGMKTPAGR----- 342
Db 296 -----ISLQYSDS-----AFLQDEL-----VGQVYVRDESRLRFYVF 326
Qy 343 PCNTADDKACAVGLGTEIKQKALAFDAS-----YSRPFRLGNTANE-----FVIGADY 391
Db 327 PTVNA-----NKQVTFSSSQDQTDQYGMKLTLSNKPMPGQWTWGLDADH 372
Qy 392 NFRSTNEQ--GRITLYARGGL-----ALNEFRSIPQVDLLANARKG-----VRGY 435
Db 373 ERFTS--NQFFPLAQASASGGLNKKIYTTGRYPSYDITNLAFLQSGVDINNLFTLNGG 431
Qy 436 SHTVATEN--LDBFGIY-----GKSTPHPADGLSLGGRLGHYKIESG----- 477
Db 432 VRYQYTENKIDDFIGVAAQRIQAGKAT--SADAFWRL--SRLRHFLENAGLLMHITEPQ 487
Qy 478 -----EGKTLHKASKTKFTGYAGAVDLDNNNSLYLSLSQLYTPQTNLDADGKLLKPR 530
Db 488 QAWLNFSGQLELDFPKY-----YGRGIYGAAYNGHLPL-----TKSVNV--SDSKL----- 532
Qy 531 QG---NOREVGKSGVMDDLARV--SFYRMDKNAAPLNPNKKTRVYALGKRVMEGV 586
Db 533 EGKVDSEYELGWR--FTGNLRTQIAAYISISDKSVVA-----NKDLTTSVDDKRIYGV 586
Qy 587 ETEISGAV-----TPKMQIHAQSYLSHSQ-----IKTAGNSRDDGIFLLMPKHS 630
Db 587 E---GAVDYLIPDQDSTGVSFNVFLKTESKVGNTQWQYDVKTASPSKA----- 631
Qy 631 ANLWTTYQVTPBELTTCGGVNAMS-----GITSSAGMHAGGYATPDMAAYRFTPKL 681
Db 632 ----TAY-----IGWAPDPWSLRVQSTTSFSDVSDAQGYKVDGTYTVDLLGSYL--PVG 679
Qy 682 KLOINADNIFNR-----HYARVGGANTFNIPGSRBTWTANLRS 721
Db 680 TUSFSIENLDFRDYTTWVGQAPLYISPGYASLYDYKGRGTF--GLNYS 729

RESULT 24
HMUR_YERPE
ID HMUR_YERPE STANDARD; PRT; 676 AA.
AC Q56989;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemin receptor precursor.
GN HMUR OR YF00283.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Yersinia.
OX NCBI_TaxID=632;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=KIM6;
RX MEDLINE=96386041; PubMed=9026634;
```

```
RA Hornung J.M., Jones H.A., Perry R.D.;
RT "The hmu locus of Yersinia pestis is essential for utilization of
RL free haemin and haem-protein complexes as iron sources.";
RN Mol. Microbiol. 20:725-739 (1996).
RC [2]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527 (2001).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; U60647; AAC64866.1; -.
CC EMBL; AJ414141; CAC89146.1; -.
CC InterPro; IPR000531; TonB boxC.
CC Pfam; PF00593; TonB boxC.1.
CC PROSITE; PS00430; TONB DEPENDENT REC_1; 1.
CC PROSITE; PS01156; TONB DEPENDENT REC_2; FALSE_NEG.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 28 POTENTIAL.
CC CHAIN 29 676 HEMIN RECEPTOR.
CC SITE 44 51 TONB BOX.
CC SITE 659 676 TONB C-TERMINAL BOX.
CC SEQUENCE 676 AA; 74230 MW; 84ED731CB914ACD3 CRC64;

Query Match 5.2%; Score 195; DB 1; Length 676;
Best Local Similarity 19.1%; Pred. No. 7.2e-06;
Matches 147; Conservative 104; Mismatches 306; Indels 214; Gaps 33;

Qy 18 LAALSSVFAAQTADLETVHIKQGR--SYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQ 75
Db 16 LAIACTPLATQAADTTTQTSSKKHSTDTMTVTATGNERSSF-----EAPM 62
Qy 76 SVSITNQOVKDRNVDFDQLARKTPGLVLSNDDGRSS---VYARGVEYSEY--NIDGL 130
Db 63 MVTVIEGNAPTSQTAATAADMLRQVPLTVTGS--GRTNQGVVNRGYGKGVLTLDGV 120
Qy 131 PAQMQS--INGTLPNLFAPDRVEVNRGPGSLFDSGEMGGIV-----NLVRRKPTKA 180
Db 121 RQGTDTGHLNSTFLDPAVKXIEIVRGPAALLYGSGALGVIAETVDAADMLQPGQNSG 180
Qy 181 PQHAAAGFGTH-----KQYKAEADVSGSLN-----SDGSVRGVMAQTVGASPRPAEK 230
Db 181 YRVYSSAATGDSHSGFGLGASAFGRDLDGLSFGTRDIGNIR-----QSNQFN---APND 232
Qy 231 NRHETFYAAADWDINDPTVLGAGLYQOORHLAPYNGLPADANNKLPSLPQHVFGADVANK 290
Db 233 ETISNVLAGTWQIDSTQSLSANLRY-----YNNSAIEPKNPQTSAPSSNTVMTNRST 285
Qy 291 FKMNSHDVFADLKHFGNGGKVGKMYRSDRSDSNYAFAGSKLGMK-----TPAGRPG 344
Db 286 IQRDAQ-----LRNPKPLDQEWLNATAQVYVSEVINEARPG-- 323
```



```

QY 345 CMTADKACAVGLGTEIKOKALAFDAS-----YSRPRGNTANEFVIGADN 392
Db 324 --SAEGREQTEGVLKERTLRFLISSPASHLLTYGTEYTKQDPTGATSPPOAK--I 379
QY 393 PRSTEGOCRTLYARGALNEFRSIPVDLIANR-KGVRSYHTVATENDEGIVG 451
Db 380 RRSGLMLQDEITL-----RDLF-VSIAGTRYDNYSSSGYADVDDPKSSRK 427
QY 452 KSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASXTKFTGVA-----GAVYDLND 503
Db 428 AISITPTDMLMFG-----SVAGAPRAPMTGMY--NDS 459
QY 504 NSLY-----LSISQLYTPTQNLADGKLKPRQNGQFVGYGYSMD-----DRINAR 552
Db 460 KHFAPIRPRLTLTYNVWPNPN-----LKPEINTEOEGFGLRPSDLMAEDLQFXY 512
QY 553 SFYRMKDKNAAAPLNNKKTRYAALGKRVMEVEFEISGAVTPKQIHAGY-----S 605
Db 513 SFYDTYAKDYIS-----TR--VDMQMTTTSVNIIDAKIWMDSKSYKTALFNMD 561
QY 606 YLHSQIKTASNSRDGIFILMPKHSANL-----WTTYQVTEBLTIGGVNAM 652
Db 562 LAYNRKGNQNTDEWLDITNPDTVTSIVDPVANGSFGVW-----IGTANNS 611
QY 653 SCGITSAGHAGGYATPDMAAYRFTPKLK--LQINADNIFNRHYARYG 700
Db 612 SVSSSTPO--AGYGVNDFVSYGQEAFFKGMTTMLAGNVFEKEYTPOG 660

RESULT 25
FEPA_ECOLI STANDARD; PRT; 746 AA.
AC P05825; P75722; P76821; P77093;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferrienterobactin receptor precursor (Enterobactin outer-membrane
receptor).
GN FEPA OR FEP OR FEUB OR B0584.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RX MEDLINE=86278160; PubMed=3015941;
RA Lundergan M.D., Kadner R.J.;
RT "Nucleotide sequence of the gene for the ferrienterochelin receptor
that interact with TonB."
RT J. Biol. Chem. 261:10797-10801(1986).
RN [2]
RX MEDLINE=9742617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [3]
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,

```

```

RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map."
RT DNA Res. 3:137-155(1996).
RN [5]
RX MEDLINE=8906678; PubMed=2974033;
RA Petris G.S., Brickman T.J., McIntosh M.A.;
RT "Transcriptional mapping and nucleotide sequence of the Escherichia
coli fepA-fes enterobactin region. Identification of a unique
iron-regulated bidirectional promoter."
RT J. Biol. Chem. 263:18857-18863(1988).
RN [6]
RX MEDLINE=9035449; PubMed=2201687;
RA Armstrong S.K., Francis C.L., McIntosh M.A.;
RT "Molecular analysis of the Escherichia coli ferric enterobactin
receptor fepA."
RT J. Biol. Chem. 265:14536-14543(1990).
RN [7]
RX MEDLINE=99101384; PubMed=9886293;
RA Buchanan S.K., Smith B.S., Venkatramani L., Xia D., Esser J.,
RA Palnitkar M., Chakraborty R., van der Helm D., Deisenhofer J.;
RT "Crystal structure of the outer membrane active transporter FepA from
Escherichia coli."
RT Nat. Struct. Biol. 6:56-63(1999).
CC -I- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
UPTAKE BY BINDING FERRIENTEROBACTIN (FE-ENT), AN IRON CHELATIN
SIBEROPHORE THAT ALLOWS E. COLI TO EXTRACT IRON FROM THE
ENVIRONMENT. FEPA ALSO ACTS AS A RECEPTOR FOR COLICINS B AND D.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -I- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M13748; AAA65994.1; -
DR EMBL, AE000163; AAC73685.1; -
DR EMBL, U82598; AAB40783.1; ALT_INIT.
DR EMBL, D90700; BAA35225.1; -
DR EMBL, J04216; AAA23756.1; -
DR PIR, A25953; ORECF.
DR PDB, 1PEP; 13-JAN-99.
DR ECODBASE, D079.0; 6TH EDITION.
DR Ecogene, EG10293; fepA.
DR InterPro, IPR000531; TonB_boxC.
DR Pfam, PF00593; TonB_boxC_1.
DR PROSITE, PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE, PS01156; TONB_DEPENDENT_REC_2; 1.
KW Transmembrane; Outer membrane; Iron transport; Transport; TonB box;
KW Signal; Receptor; 3D-structure; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 746
FT SITE 34 41
FT SITE 729 746
FT CONFLICT 152 152
FT CONFLICT 403 403
SQ SEQUENCE 746 AA; 82107 MW; 09348AAB1C29710A CRC64;

Query Match
Best Local Similarity 19.1%; Pred. No. 1,1e-05;
Matches 153; Conservative 108; Mismatches 280; Indels 261; Gaps 35;
QY 77 VSIITNQGVKDRNV-DTFDLARKTPGLRVLSN-----DDGRSSVARGYSEYSEYND 128

```

```
Db 50 VSTIADIRKPNVARDVSKIIRTWPGVNLGTSGQGNRRQIDIRGMGPENTLIIID 109
Qy 129 GLPAQMOS-----INGTLNLPFAFDREVVRMGPSGLFDSSGEMGIVNLV 173
Db 110 GKFVSSRNVROGWRGDRTRGDTSWPPM--IERIEVLGRFAAARYGNGAAGVWNII 167
Qy 174 RKRPTKAFQGHAAAGFT--HKQ-----YKAEADVSGSLNSDGSVGRVMAQTVGASPPA 227
Db 168 TKKGSGEWHGSDWAYFNAPEHKEGATKRTNFSLTGTLGDEFSPR----- 212
Qy 228 EKNRRHETFYAAD-----WDINPDTVLGAGLYQOQRLHAPYNGLPADANNKLPSPQ 281
Db 213 -----LYGNLDKTDQADWDIN-----QHQSRAGATYA---TTLTLAGREG 249
Qy 282 VFGAD-----WNKFKQNSHDVDFADLKHYFGNGGKYGVGMRYSDRDADSNYAFAGSKL 334
Db 250 V-INKDINGVRWD-----FAPLQSELEAGYSRQGNLYAGDTQNTNSDSYTRSKY 299
Qy 335 GSKT-----PAGRPGCNTA---DDKACAVG 356
Db 300 GDETNRLYRQNYALTWNGWDGVTTSNWVQYEHFTRNSRIPEGLAGGTGKFNKATQDF 359
Qy 357 LGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRR-----STNEQGRITLYARGGLA 412
Db 360 VDIIDDDVNLHSEVNLPIDFLVNQI---LTGTENWQRMKDLSSNTQALQ----- 408
Qy 413 LNEFRSIPQVDLIANARKGVGRVYSHTVAT--ENLDEFYIGYKSTPHPADGLSLIGGRLG 470
Db 409 -----TWTGAIDGVSTTDRSPYSKAEIISLFAENNMELTDTSTVTPGLRFD 455
Qy 471 HYKIES-----GEGKTLHKASKTKPTGYAGAVYDLDNNSLYLSLSQLYTP- 516
Db 456 HHSIVGNWNWSPALNISQGLGDDFTLKMGTARAYK--APSLYQTNPNYILYSGQGCYASA 513
Qy 517 -----QTNLDADGKLLKPRQGNQFVGKYSYMDRLNARVSFYRMKDNAAAPLNPNNK 571
Db 514 GCVYQGNDD-----LKAETSINKIEIGUE--FKRQGLWAGVTWFRNDRNKI----- 558
Qy 572 KTRYAALG-----KRVMEGVETEISGAV--TPKQIHAQYSYLVHSQIKTA 614
Db 559 EAGYVAVGQNAVGTDLQWQNVKPAVVEGLESLNVPVSETVMWNINITY-MLKSENKTT 617
Qy 615 SNRRDGGFLMLPKHSANLWTTYYQVTPBELTIGGGVYNAMSGIT-----SSAGMIAG 664
Db 618 GDR-----LSIIPEYTLNSTLSQAREDSM-----QTTFTWYKQOPKKNYKQGPAP 666
Qy 665 G-----YATFDAMAAYRFTPKLQIADNADNIFNRHY-----YARVG 700
Db 667 GPETKEISPYISVIGLSATWDVTQVNSLTGCVGNLFDKRLWRAGNAQTGDLAGANIAGA 726
Qy 701 GANTFNIGSERTWNLIRYSF 722
Db 727 GAYTNEFG--RTWYMSVNTHF 746

RESULT 26
ID_RHTA_RHIME STANDARD; PRT; 746 AA.
AC O923Q5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rhizobactin receptor precursor (TonB-dependent siderophore receptor rhtA).
DE RHTA OR Ral265 OR Sma2414.
GN Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymba (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=21172875; PubMed=11274118;
RA Lynch D., O'Brien J., Welch T., Clarke P., Cuiv P.O., Crosa J.H.,
RA O'Connell M.;
RT "Genetic organization of the region encoding regulation, biosynthesis,
RT and transport of rhizobactin 1021, a siderophore produced by
RT Sinorhizobium meliloti.";
RL J. Bacteriol. 183:2576-2585 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federpiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymba megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
CC -!- FUNCTION: RECEPTOR FOR THE SIDEROPHORE RHIZOBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/submit/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF110737; AAD09419.1; -
DR EMBL; AE007312; AAK65923.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Plasmid; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 746 RHIZOBACTIN RECEPTOR.
FT SITE 40 47 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80633 MW; 16AE44A4025D5B23 CRC64;

Query Match 5.1%; Score 193.5; DB 1; Length 746;
Best Local Similarity 18.9%; Pred. No. 1.le-05;
Matches 163; Conservative 116; Mismatches 308; Indels 277; Gaps 38;

Qy 10 INMTAATVLAALSSSVFAAQAD-----LETHIKGQSYNAIVTEKNGDYGSFATV 62
Db 9 ISFCVFWVVGFGTGAQAQEPANQSEAVTSLEEIVVTGGRSAQOI-----SEIARTI 60
Qy 63 ----GTXIPASLRIPQSVSIITNQOVKDRNVDFD-----QLARKTGLRLVS 107
Db 61 YVVDSDQIQAEAR-----SGTKLQQLGETIPSPASDGARTSFQGNLRGRPLILV- 113
Qy 108 NDDGRSSVYARGYEYSEYNIDGLPAQMQSINGTLPNLPFAFDREVVRMGPSGLFDSSGEMG 167
Db 114 --DGVSMNSAR-----SLSRQFDAIDP-----FNIERVEVLGATAIYGNNA-TG 155
Qy 168 GIVNLVRKPTKAFQG-HA--AAGFGTHKQYKABADVSGSLNSDGSVGRVMAQTVGASP 224
Db 156 GIINIITKKGKDAEPGLHAEVTVGMG-----SGFAGSQDFDRNAAGAVTYNSENWDA 207
Qy 225 RPAENRHEHETFYAADWDINPDTVLGAGLYQOQRLHAPYNGLPADANNKLPSPQH--- 281
Db 208 RLSTAGNRGTAFYDGSGLTLLIPDITQTSTAFNERIDLMSGISGYQIDDDRVEFGSPDS 267
Qy 282 -----VFVG-----ADMNKPKONS-----HDVFAD----- 301
Db 268 QKSDSYGLIYGPFFPAALADPSLFETRSGYESDFNPQTRRSMLNVTYTNDNVFGQQLLQG 327
```



```

Db      612 RGLNLNVEFEPVFGTFLASATWLDPKLTAAASANNNG-----NDAGVANYQ-----L 660
Qy      644 TIGG-----GVNMSGITSSAGNAG-----GYATPDAMAAYRTPK---LKIQ 684
Db      661 VEGGEVDIVVEGLTATGTIVVRSQGYANBANTLKXKPTMRDLGVRITPMMDTSLTMR 720
Qy      685 INADINFNHYARVGANTFNPISERTWTANLRYSF 722
Db      721 ANIENTNERWESVEDSGTYIQGDPRALKLSVMDF 758

RESULT 29
TBPL_NEIMO STANDARD; PRT; 915 AA.
AC 001966;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=92394880; PubMed=1325963;
RA Cornelissen C.N., Bismar G.D., Tsai J., Paruchuri D.K.,
RA Thompson S.A., Spirling P.F.;
RT "Gonococcal transferrin-binding protein 1 is required for transferrin
RT utilization and is homologous to TonB-dependent outer membrane
RT receptors.";
RL J. Bacteriol. 174:5788-5797(1992).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC CC
DR EMBL, M96731; AAA25503.1; -
DR PIR, A43335; A43335.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KM Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 915 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB_BOX.
FT SITE 898 915 TONB_C-TERMINAL_BOX.
SQ SEQUENCE 915 AA, 102213 MW, 657CFT4B10104222F CQC64;

Query Match 4.6%; Score 172; DB 1; Length 915;
Best Local Similarity 21.4%; Pred. No. 0.00042;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

```

```

Db      101 IRMDNRVSLTVDDL-AQIQSYTAQALGTRTAGSSAINIEIEENKAYAEISGSNS 159
Qy      159 LFDSSGEMGIVLVKRPPTKAFQGHAAAG-FCITHQYAEADVSGLSNDSGVRGRVAA 217
Db      160 VEGGSALAGSV-----AFQRTADVDIGEGHOMQIGSTAYSGKRGITQSLALA 210
Qy      218 QTVGASPRAEKNNRHETRYAAMDINDPVLGAGLYVQORHAY----- 264
Db      211 GRIGAEALLIRGRRA-----GEIRAEAAAGR-VQSFRLAVDDGSKYAFIVEE 262
Qy      265 ---NGLPADANNTPLPQHVFPVGADW-----NKRKM--SHDV----- 298
Db      263 ECKNG-----GHEKCKANPKKDVGEDKQTVSTRTDTGPNRFLADLSYESRSLRPPGF 318
Qy      299 -PADLGHYFGNGGCVKGRYSRDA-----DSNYAFAGSKLTKMPAGR---P 343
Db      319 RENKSHYTG-GILERTQOTFTRDTVPALTKAVFDNMQOAGSLRNGKYAGNHYG 377
Qy      344 GCNTADDKACAVG--LGTEIKOKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEOG 401
Db      378 GLFTSGENNAAPVGAEGTGVFDETHTKSRYGLEYVYTNADK--TWADYARLS----- 429
Qy      402 KTTLYARGLAL-NEFRSIPQVDLIANARKGVRGYSHTVATENLDEFG---LYGKSTH 456
Db      430 ---YDRQIGLDNHFO---QTHCSAD---GSDKYCRPADKPFYSYKSDRYVGBS--- 476
Qy      457 PADGLSLIGGRLGHYKHSIGEGKTLHKASKTF-----TYGAAGVYLDLNDN 504
Db      477 -----HKLQAAFPKSFDTAKIRHNLNVNLGYDRFGSNRHOD 514
Qy      505 SYLISLSQUTPOTNLADAKLLKP--ROGNOFEVYKSGSYMDRLNANVSFFRMQKXA 562
Db      515 YYQGSANRAVSLKTPPQNNKGKTSPPGKREKNPYWSI-----GRNVVTRQCLFGNT 568
Qy      563 AAPLPNN--NKTIRVALAKRVMEGVETEISGAVTPKW-QIHAGYSY---LHS---QIK 612
Db      569 YVDCSTRSINGKSYAAVDNRLG-----RMADVAGALRYDRSTHSDGSYS 617
Qy      613 TASN---SPDDGFILMPGHSANLMTTQVTPELTIGGVNMSGITSSAGNAGYATF 669
Db      618 TGHRTLSNAGIVL--KPADWLDLTYR-----TSTGRFLPSFA-- 654
Qy      670 DMAAAYRFTPKLK-LQINDNIFNR 693
Db      655 -EMYGWRSQDKIKAVKIDPEKSPNK 678

RESULT 30
TB12_NEIMO STANDARD; PRT; 908 AA.
AC 006987;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=491;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-42.
RC STRAIN=CCUG 37603 / B16B6 / Serogroup B / Serotype 2a;
RX MEDLINE=93345825; PubMed=8344530;
RA Legrain M., Mazatrin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
RA Jacobs E., Schryvers A.B.;
RT "Cloning and characterization of Neisseria meningitidis genes
RT encoding the transferrin-binding proteins Tbp1 and Tbp2.";
RL Gene 110:73-80(1993).
RN [2]
RP SEQUENCE OF 25-45.
RC STRAIN=CCUG 37603 / B16B6 / Serogroup B / Serotype 2a;
RX MEDLINE=93307625; PubMed=8319886;

```

RA Griffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,
 RA Borriello S.P., Holland J., Parsons T., Williams P.;
 RT "Antigenic relationships of transferrin-binding proteins from
 RT *Neisseria meningitidis*, *N. gonorrhoeae* and *Haemophilus influenzae*:
 RT cross-reactivity of antibodies to NH2-terminal peptides.";
 RL FEMS Microbiol. Lett. 109:85-91(1993).
 CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
 CC TRANSFERRIN UTILIZATION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: BY IRON STARVATION.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z15129; CAA78831.1; -;
 DR PIR; JN0819; JN0819.
 DR PIR; S33154; S33154.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Receptor; Signal; TonB box.
 FT SIGNAL 1 24
 FT CHAIN 25 908 TRANSFERRIN-BINDING PROTEIN 1.
 FT SITE 38 45 TONB BOX.
 FT SITE 891 908 TONB C-TERMINAL BOX.
 SQ SEQUENCE 908 AA; 101583 MW; FE2FF4974CAC3C31 CRC64;

Query Match 4.4%; Score 167.5; DB 1; Length 908;
 Best Local Similarity 18.2%; Pred. No. 0.00085;
 Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

QY 3 QPMVSFRINMTAATVLAALSSVFA-----AQTADLETVHIKQSYNAIVTEKNGDY 55
 DB 2 QOHLFRNLCLSLMTAL--PYAENVQAEQAEKQDITQVAKKQKTRDNEVTG-- 57
 QY 56 SSFATVGTGTPASIRETPQSVITNQVKORNDVDFDQARKTPGLRVLSNDDGRSV 115
 DB 58 -----LGLVKSSDTLSKEQVLN-----IRDLTRYDPGIAVVEQGRGSSG 98
 QY 116 YA-RGVEYSEYN--IDGLPAOMQS-----INGTLP-----NLFAPDR---VEVMRGP 156
 DB 99 YSIRGMKRNRLTVDGV-SQIQSYTAQALGTTAGSSGAINIEYENVKAVEISKGS 157
 QY 157 SGLPDSGSGMGIVN-----SGLPDSGSGMGIVN-----SGLPDSGSGMGIVN 171
 DB 158 NSSEYNGCALAGSVAFTQKTAADIIGEGKQWGIQSTAYSGKDHALTQSLALAGSGGAE 217
 QY 172 ---LVKRPRTKAPQHAAGFGTH-----KQVK-----KQVK-----ABADVSG 203
 DB 218 ALLIYTKRRGRRIHAHKDAGKQVQSFNRLVDEKKEGSOYRYFIVEECHNGYAAACKN 277
 QY 204 SINSDGSV---RGRYMAQTVGASPR---PAE-----KNNRH-----KNNRH 233
 DB 278 KUKEDASVKDERKTSTQDYTGSNRLANPLEYGSQSWLFRPGWHLNDRHYVGAVLERTQ 337
 QY 234 ETF-----YAAADWDINP-----DTVLGAGY-----LY 256
 DB 338 QTFDTRDTPVPAFTSE-DYVPSGLKGLGKYSQNDKAEERLFVQGGSTLQGIYGTGVFY 396
 QY 257 QORHLAPYNGLPADANKLPSPHQHVFGADWNKFKMNSHDVFDLKHVFGNGGKGVKM 316
 DB 397 DERHTKNRYGV-----EYVYNAD-----KDTWAD-----YARLSYDROGI 432
 QY 317 RYSDRDAQSNYAFAGSKLGMKTPACRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 376
 DB 433 DLNRLQOHTCHSDGSDKNCRPDGNKPYFYKSDMI-----YEESRNLFOAVFKAP 485

QY 377 RLGNANTAEFVIGADYNPRSTNEQGRRTLYARGGLALNEFRSIPQVDLIANAR-----KGV 432
 DB 486 DTAKIRHNLSINLGYDRFKS--QLSHSDYYILQ-----NAVQAYDLITPKKPPFPNGS 535
 QY 433 RYSHSTVA---TENLDEFGIYKSTFHPADGLSLIGG-----RLGHYKIESGEG 479
 DB 536 KDNPYRVSIGTKTNTSPICRFGNNTYTDCPTRNIGNGYAAVODNVRLGRW-ADVAG 594
 QY 480 -----KTLHKASKTKFTG-----YAGAV-----YDLNDNNSLYL---SLSOLYTPQT 518
 DB 595 IRYDRSTHSEDKSVSTGTHRNLWNAGVVLKPTWMDLTYRSTGPRLPSPFAEYGHRA 654
 QY 519 NLDADGKLLKPRQGNQFEVG--YKG-----SYMDRLNARVSF-YMKDKN-----AA 563
 DB 655 GESLTKLDLKPESKFNREAGIVFKGDFGNLEASYFNNAIRDLIAFGVETRTQNGQTSASG 714
 QY 564 APLPNPKKTYAA--LGRVMEGVETEISGAVPKW-----QIHAGYSYLHSGQIKTAS 615
 DB 715 DPGYRNAQNAIRAGINILGKIDWHGV-----WGLPDGLYSTLAYNRKVKDA- 762
 QY 616 NSRDDGIFL-----LMPK-----HSANLW-----TTYQ-----VTPELTIG 646
 DB 763 DIRADRTFVTSYLPDAVQPSRYVLGLGVDHPDGIWINTMTFTSKAKSVDELGSQALLN 822
 QY 647 GGVNAMSGITSSAGHAGGYATFDMAAAYRFTPKLKIQINADNIFNRHY----- 695
 DB 823 GNANA-----KKAASRRTRPWTVDVSGYINIKKHLTLRAGVYNLLNRYVTWENVRTAG 878
 QY 696 -----YARVGGAFTNIPGSERTWTANLRY 720
 DB 879 GAVNQHKNVGVYNYRPAAPGRNYTFSELMKF 908

RESULT 31
 HXC2_HAEIN
 ID HXC2_HAEIN STANDARD; PRT; 725 AA.
 AC P45357;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Heme/hemopexin utilization protein C precursor.
 GN HXUC.
 OS *Haemophilus influenzae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC *Haemophilus*.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DL42 / Serotype B;
 RX MEDLINE=95270579; PubMed=7751272;
 RA Cope L.D., Yegor R., Mueller-Eberhard U., Hansen E.J.;
 RT "A gene cluster involved in the utilization of both free heme and
 heme:hemopexin by *Haemophilus influenzae* type b.";
 RL J. Bacteriol. 177:2644-2653(1995).
 CC -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
 CC CONCENTRATIONS.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE D OF
 CC H.INFLUENZAE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U09840; AA87059.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.

DR PROSITE, PS00430; TONB DEPENDENT REC_1; FALSE_NEG.
 DR PROSITE, PS01156; TONB DEPENDENT REC_2; 1.
 KW Outer membrane; Transport; Tonb box; Signal; Receptor.
 FT SIGNAL 1 19 OR 21 (POTENTIAL).
 FT CHAIN 20 725 HEME/HEMOPROTEIN UTILIZATION PROTEIN C.
 SQ SEQUENCE 725 AA; 80838 MW; FCT8865020CBCE CR664;

Query Match 4.4%; Score 165.5; DB 1; Length 725;
 Best Local Similarity 20.1%; Pred. No. 0.0085;
 Matches 167; Conservative 115; Mismatches 331; Indels 217; Gaps 43;

QY 5 MGVPRNMTAAVYLAALSSVFAAQTADLTHIKQR--SYNAVTEKNQDYSFAVTV 62
 DB 1 MEEKSLIAIATTLVLANA---LAQSVELDSIVIAIRDPSPRFAVPEKQSK--DSLISKO 56
 QY 63 GKIPASLEIPQSVSIITNOQVKNRVDTF--FDLARKTPELRYLSDN-----DGRS 113
 DB 57 AIVSAALADIDP-----NVDIRGSSRSIAQK--PIKRGSLDRVVQVIVDGR 101
 QY 114 SVYARGYSEYNIDGLPAQOMOSINGTLNPLFAFDREVMWGRSGPLFDSGEMGIV--- 170
 DB 102 QNFDLAHRGSYF---LPMSL-----IQIEIVIKGPPSSSLMGSALGVVAMR 145
 QY 171 ----NLVR---KRPYKAFQGHAAAGFTHKQY-----KAEADVSGSLNSDGVGRV 215
 DB 146 TENALDLKKNPKFGYKIQGYQTANNLSERDVSFPAANDKFDVLISGFYNADNL----- 201
 QY 216 MAQTGASRPAPKKNRHETFY-----AAADNDINPDVYLGAGYLYQO--RHLPYNGLP 268
 DB 202 -----RTGKGNKLNATYAKQFGGLAKFGQINDANKEVLSHETRRKQTPANN-- 250
 QY 269 ADANNKLP-----LPOHVEFGADMNFKKNSHDVFDLKHYPNGGCGYKGMESDR 321
 DB 251 -EVENMLTEQITDQIREFHKKPNKSSPPKAKPSQEEFYSVKTRFSGVSF--LTDQDIPDQ 308
 QY 322 DADSNAFAGSKLGMKTPAGRPGCNTADKACAVGLGTEIKQKALAPDASYSPPRLGNT 361
 DB 309 STVFNYVYLLPDPNYLNTHTALYNNKTIKEQRKVS--GVKDQDTLTL-----TRGILRNS 361
 QY 382 AN-----EFTYGADYNFRSTNEGRTTLVARGGLANEPISIQOVULINARQVAGYSH 437
 DB 362 SELSHISFYGVVDYMDKIRTERGTNNKDA-----QFPADY-----NAN-----SN 403
 QY 438 TVATENLDEFGYKSGSTFPADGLSLIGGRLGHYKLESCEGKTL-----HKASKTK--- 489
 DB 404 TTGVYLIATIIPLFEGK-----LILSPSVRYKHVDTSS---KTVYKXKNHLSPATKLTW 453
 QY 490 -----FTGVAGAVYDLNDSNLSYLSQLYT-----PQT----- 518
 DB 454 IVTNMLDPTAKYNEAFRAPSMMQERFVSGSHFSTILGRNEINKFVANPNLRPTAKNKEI 513
 QY 519 --NLDDGLKLRQGNQF--EYGYKGSYMDDLNARVSYRRKDKGAAP-----LNP 568
 DB 514 TNLHLSLDFP---KQDDEKRIETVFRNDVKDFINKLIFENDAKTNTNASSAGAGAGANP 570
 QY 569 NN-----KTRRYALGRVWEGVETEISGATPEKQIHAGSYLSHQIKTASNRDDGIFL 624
 DB 571 NGALLPTKQYQYNTNARLSGIELQ--AQYVTEHLTLFTYG-----STGKXKXDSGEAL 623
 QY 625 L-MPKHSANILMTTYQ--VTDELITGGV--AMSGITSSAGMHAGYATFDMAAAYRTP 679
 DB 624 SNIASKIGVGVYALVKDKFTVGAATVTHYAQRVRPKDHSVTPSYLILDLDAAT--AP 661
 QY 680 -----KLTQINADNIFNRHYARVAGANTFNI--FGSEETWTANLRYSF 722
 DB 682 LKGEWKNLRLDPALENLFDKRYQF-----AFSLMEGTGNAKISAVYSF 725

RESULT 32
 HPUB NEIMA
 ID HPUB NEIMA STANDARD; PRT; 810 AA.
 AC Q9JWA2;
 DT 16-OCT-2001. (rel. 40, Created)

DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Hemoglobin-haptoglobin utilization protein B precursor.
 GN HPUB OR NMA0474.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parthill J., Achtmann M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
 RA Jagels K., Leather S., Moule S., Murgall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491".
 RL Nature 404:502-506 (2000).
 CC - FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
 HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE (BY
 SIMILARITY).
 CC - SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC - SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AL162753; CAB83769.1; --
 DR InterPro; IPR000531; Tonb_BoxC.
 DR Pfam; PF00593; Tonb_BoxC; 1.
 DR PROSITE; PS00430; TONB DEPENDENT REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB DEPENDENT REC_2; 1.
 KW Outer membrane; Transport; Tonb box; Signal; Receptor;
 FT Complete proteome.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 810 HEMOGLOBIN-HAPTOGLOBIN UTILIZATION
 FT SITE 793 810 TONB C-TERMINAL BOX.
 FT SEQUENCE 810 AA; 90570 MW; D38DEIDCA3CASA6E CR664;

Query Match 4.3%; Score 162.5; DB 1; Length 810;
 Best Local Similarity 19.2%; Pred. No. 0.0016;
 Matches 168; Conservative 109; Mismatches 306; Indels 293; Gaps 45;

QY 55 YSSFAVTGKTIPLASIREI---POSVSIITNOQVKNRVDTF-----DQLARKTGLKRV 105
 DB 18 FPAFAADPAPOSQOTLNELTVGTHTOKLGEKIRKRTDKLVNDEHDLVRYDDGISV 77
 QY 106 LNSDDRSSVYA--RGVEYS--YNIDGLPAQ-----MOSINGTLPLMLFA----- 146
 DB 78 VEGGRAGSNGFTIRGVKQKRVAINVDGL--AQAESRSSEAFQELFGAYGNFANRNTSBE 136
 QY 147 -FDRVEVMRGPGLFDSGEMGIVNLVRK-----RPTRAFQ 183
 DB 137 NFSEVITITGADSLKSGSGLAGAVVYQTKASDVYSEDKPYHLGKGSVGNKSGKFS 196
 QY 184 HAAAG-----FGTH--KQYKAEADVSGSLNSDGSYRGVMAQTGASRPAP 228
 DB 197 ITAAGRLFGIDLALLVYTRRFGEKTKNRSTEGDI--EIKNDGVY-----YNPLDTG 244
 QY 229 KNNRHETFYAAA-----DMDINPDVYLGAGYLYQOHLAPY---NGLPADANNKLP 276
 DB 245 GPSKYLITVATGVANSQDPQDEK--VAKSTLFLKGIYFNQNRIGMIFEDSRIDRFTNELS 303
 QY 277 SLFQHVFGADMNKFKGNSHDVF-----ADLKHYFGNGGYGKVGKRYSDR----- 321

Db 304 NLWTGTTTSAATGDIR-HRQDVSYRRRSVEYKNELEHCPWDSLKLRYDKQRIQDMNTWTW 362

Qy 322 DADSNYAFAGSKLGMKTPAGRPCCNTADDKACAVGLGTEIKOKALAFD-----ASYSRPF 376

Db 363 DIPKNYKRGINGEVVHSPHRIQNTAQ-----WTADFEKQLDFSKAVWAAQYGLGG 414

Qy 377 RLGNATANEFVIGADYNRFRSTNEQGRRTLYARGGLALNEFRSIPQVDLIANARKGVRGYS 436

Db 415 GKGDAN-----SDYSYF-----AKLYDPKILASNOAKI----- 443

Qy 437 HVAATENLDEFGI-YGKSTFHPADGLSLIGGRLGHYKIESG----- 477

Db 444 -TMLNRSKYFAWNNAFH-----LGGN--DRFRNLNAGIRYDKNSSAKDDPKYTT 493

Qy 478 --EGKTLHKASKTKETGYA-GAVYDLNNDNSLYLSLSQLYT-----PQT 518

Db 494 ATRGOIPIHLGSEAHAGFSYGVGDFWRFTKHLHL-LAKYSTGFRAPTSDETWLLFPHPDF 552

Qy 519 NUDAGKLLKPRQNOFEVYKGS-----YM-----DDRLNARV 552

Db 553 YLKANPN-LKAERKANWELGLAGSKAGNFKLSGFKTKYRDFIELTYMGVSDDDKNR- 610

Qy 553 SFYRMKDNA--AAPLNPNKKTRVAALGKRVMEGVETEISGAVTPKQIHAQ--YSYLH 608

Db 611 -YAPLSDGTALVSSPWQN--QNRGAANWVGIEFNGTWNLDIGLPK-GLHTGLNVSYIK 666

Qy 609 SQIKTASNRDDGIFLLMPKHSANLWTTVQVTPELTIG-----GGVNAMEGITSSA-- 659

Db 667 GK-ATQNGKETPINALSP-----WTAV-----YSLGYDAPSKRWGINAYATRTAAKPK 714

Qy 660 -----GMHAGGYATFDMAAYRFTPKLKQLQINADNIFNRHYA----- 697

Db 715 SDTVHSNDLNNPWPYAKHSKAYTLFDLSAYLNIQKQVTLRAAAYNITNKQYTTWESLRS 774

Qy 698 -----RV-----GGANTENIPGSERTWTANLRY 720

Db 775 IREFTGVNRVDNKTAGIORFTSPGRSNTIEAKF 810

RESULT 33

ID TB11_NEIMB STANDARD; PRT; 911 AA.

AC Q09056;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transferrin-binding protein 1 precursor.

GN TBPL.

OS Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=491;

RA (1)

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=CCUG 37608 / M982 / Serogroup B / Serotype 9;

RX MEDLINE=93345625; PubMed=8344530;

RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J., Jacobs E., Schryvers A.B.;

RT "Cloning and characterization of Neisseria meningitidis genes encoding the transferrin-binding proteins Tbp1 and Tbp2.";

RL Gene 130:73-80(1993).

CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR TRANSFERRIN UTILIZATION.

CC -!- SUBCELLULAR LOCATION: Outer membrane.

CC -!- INDUCTION: BY IRON STARVATION.

CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; Z15130; CAA78833.1; -.

DR PIR; JN0821; JN0821.

DR PIR; S33156; S33156.

DR InterPro; IPR000531; TonB_boxC.

DR Pfam; PF00593; TonB_boxC_1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

KW Outer membrane; Receptor; Signal; TonB box.

FT SIGNAL 1 24

FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.

FT SITE 38 45 TONB BOX.

FT SITE 894 911 TONB C-TERMINAL BOX.

SQ SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773E6 CRC64;

Query Match 4.3%; Score 161; DB 1; Length 911;

Best Local Similarity 18.3%; Pred. No. 0.0024;

Matches 178; Conservative 133; Mismatches 346; Indels 314; Gaps 49;

Qy 3 QFMSVFRIN--MTAATVLAALSSVFA--AQTADLETVHIKQRSYNAIVTEKNGDYSS 57

Db 2 QQQHLFRNLNCLSLMTALPAYAENVQAQQAQKQLDTIQVKAQKQKTRRDNEVTG---- 57

Qy 58 FAVTVGTKIPASLREIPQSVSIITNOQVKDRNVDTFQOLARKTLCRLVLSNDDGSSVYA 117

Db 58 -----LGLVKTADTLSKEQVLD-----IRDLTRYDPIAVVEQGRGASSYS 100

Qy 118 -RGVEYSEYN--IDGLPAQMOS-----INGTLP-----NLPAFDR---VEVMRGPSP 158

Db 101 IRGMDKNRVSITVDGL-AQIQSYTAQAALGTRTAGSSGAINEIYENVKAVEISKGSNS 159

Qy 159 LFDSSGEGGVIVNLRKPTKAFQGHAAAG-FGTHQYKAEADVSGSLNSGSGVRGYMA 217

Db 160 VEQSGSALAGSV-----AFQTKTADDDVIGEGROWGIQSKTAYSGKNRGLTQSIALA 210

Qy 218 QTVGASPR-----PAEKNRHEFP----- 236

Db 211 GRIGGAELLIHTGRRAGEIRAHEDAGRGVQSFNRLVEPDESSEYAYFIVEDECEGKRYE 270

Qy 237 -----YAAADWDINPDTVLGAGLYQ-QR 259

Db 271 TCKSKPKDVVGKDERQTVSTRDYTGPNRFLADPLSYESRSLFRP-----GFRFENKR 324

Qy 260 HLAPYNG-----LPA-----DANKK-LPSLPQH----- 281

Db 325 H---YIGGLEHTQQTFTDRDMTPVFLTKAVFDANSKQAGSLPQNGKYAGNHKYGGLFT 381

Qy 282 -----VFVGADWNKFKMNSHDVFDLKHFGNGGKVGKGMRY-----SDRADSNYA---F 329

Db 382 NGENGALVGAEGYT-----GVFYDETH-----TKSRYGLEVYVTNADKDTWADYARLSY 430

Qy 330 AGSKLGMKTPAGRPCNT-ADDKACAVGLGTEIK-----OKALAFDASYSRPPRL 378

Db 431 DRQIGLDNHFQQTHCSADGSDKYCRPSADKPFYSYKSDRVIYGESHRLLOAFAFKSFD 490

Qy 379 QNTANEFVIGADYNRFRSTNEQGRRTLYARGGLALNEFRSIPQVDLIANARKGVRGYSHT 438

Db 491 AKIRHNLNVNLGDFRFDPS-NLRHQDYVYQHANRAYSS--KTPPKTANPNDGKS-KPYWVS 546

Qy 439 VATENL--DEFGIYKSTFHPADGLSLIGG-----RLGHYKIESGEG-----KTL 482

Db 547 IGGNVVTGQICLFGNNNTYDCTPRRSINGKSYAAVRDNVRLGRW-ADVAGLRYDYRST 605

Qy 483 HKASKTKFTG-----YAGAV-----YDLNNDNSLYL-----SLSQLYTPQTNLDADGKL 526

Db 606 HSDGGSVSTGTHRTLSWNAGIVLKPADWLDLTYSCTGFRLPSPFAEMYGWGSQSKAVK 665

Qy 527 LKPRQGNQFEVG--YKG-----SYMD---DRLNARVSFYRMKQKNAAAPLNPN 569

Db 666 IDPKSFPNKEAGIVFKGDFGNGLEASWFNNAYRDLIVRGYEAQIKNGKEAKGDPAYLNAQ 725

Qy 570 NKK-TRYAALGKRVMEGVETEISGAVTPK-WQIHAGYSYLH-SQIKTASNSRDDGIFL-- 624

Db 726 SAATGGINIGKIDNMGVNDK-----PEGWYTFANRVHVEDIKRADRTDIQSHLFD 780

Qy 625 -LMP-----KISANLWTTYQVTPBLT--IGGG--VNAMSGITSSAGMHAAG 665

Db 781 AIGPSRYVWGLGVDPEGKGVNGLTYSKAKETELLSRALLNGSRNTKATARTRP 840

Qy 666 YATFDMAAYRFLPKLQIADNINFRHY-----YARVGANTENIPG 709

Db 841 WYVVDVSGYTTIKHFPLRAGVYNLNLRYVTWENVRQTAGAGVAMQKGVNRYAPAG 900

Qy 710 SERTWTANLRY 720

Db 901 RNTFSLEMKF 911

RESULT 34

Y262_HAEIN STANDARD; PRT; 723 AA.

AC P4600;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable tonb-dependent receptor HI0262 precursor.

GN HI0262.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OX Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RD / KM20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Keitelavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D., Utechtback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M., RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RT Science 269:496-512(1995).

RL [2]

RN IDENTIFICATION BY MASS SPECTROMETRY.

RX MEDLINE=20137488; PubMed=10675023;

RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B., Gray C., Fountoulakis M.;

RT "Two-dimensional map of the proteome of Haemophilus influenzae.";

CC Electrophoresis 21:411-429(2000).

CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.

CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).

CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL, U32712; AAC1927.1; -

DR TIGR, HI0262; -

DR InterPro, IPR000531; Tonb_boxc.

DR Pfam, PF00593; Tonb_boxc.1.

DR PROSITE, PS00430; TONB DEPENDENT REC 1; FALSE_NEG.

DR PROSITE, PS01156; TONB DEPENDENT REC 2; 1.

KW Outer membrane; Receptor; Signal; Tonb box; Complete proteome.

FT SIGNAL 1 21

FT CHAIN 22 723

FT SITE 706 723

TONB C-TERMINAL BOX.

SQ SEQUENCE 723 AA; 80775 MW; 91EB3AB0FE2A984 CRC64;

Query Match 4.2%; Score 158.5; DB 1; Length 723;

Best local similarity 19.9%; Pred. No. 0.0026;

Matches 170; Conservative 114; Mismatches 301; Indels 271; Gaps 41;

Qy 5 MSVFRIMTAATVLAALSSVFAAQTADLETTHIKQOR--SYNAIVTEKNGDYSSFAVTV 62

Db 1 MRPSKLSIALITTLTVLANA---LAQSVELDSINVIAITRDPSPRAYPEKQSK-DLSLSKQ 56

Qy 63 GTIIPASLREIPQSVSLITNQVQKDRNVD---FDQARTPELRVLSND-----DGRS 113

Db 57 ATSVADALSDIP-----NVDVGGSSRSIAQK-PNIRGLSDNRVVQVTDGVR 101

Qy 114 SVYARGYESEYVINDGLPAMQOSINGTLPLVFAFDRVEVNRGSGGLFDSGEMGAY--- 170

Db 102 QNFDLAHRGSYF---LPMSL-----IQEIEVYKGSSSLSMGALAGGVAMR 145

Qy 171 ----NLVR---KRPTKAFQGHAAAGFGTHKQY-----KABADVGSLSNDGVRGRV 215

Db 146 TPNALDLKKNDRKFGVKIRQYQTANNLSEKQSVFAANDKFDVLISGFYNNADNL---- 201

Qy 216 MAQTVGASPPPAKNNRHETFY-----AAADWDINDPTVLGAGYLYQQRHLAPYNGLP 269

Db 202 -----RTGKNKLNNTAATYRQFGGLAKFGQINDANRV-----ELSHRETRFKQTA 246

Qy 270 DANNKLPSPLPQHVFGADVWKKFQNSHDV-----PADLKHYPNGGYGKV 314

Db 247 PSNNEVENELTNEQITDOIKKFHGQKDDLPTPTQSPSPRSSEFYKVKTRLSVST-LT 305

Qy 315 GMRYSRDRDASNYAFAGSKLGKMTTPAGRCPCNTADPKACAVGIGTEIKORALAFDASYSR 374

Db 306 DQIIPQSVTFNYVYLPNDPNVYALNTALVYNNKTIEKQKVS-GVMDQTKLT-----TR 358

Qy 375 PFLGLGTAN---EPYIGADYVFRSTNEQRTTLYARGLALNERSITQVULINARK 430

Db 359 GINLRSSSELSHISFYGVADVMDKIRTERGTNGSDAK-----FRADPY-----NAN- 405

Qy 431 GVGVSHTVATENLDFEYIGKSTFHPADQLSLIGGRIGHYKIESGEKTL-----HKA 485

Db 406 ----SNTTGVYLIANIPLFEK-----LVSPSVRDIHYDTSS---KTVAYKDNHLS 450

Qy 486 SKTKFTG-----YAGAVYDLN-----DNNSLYLSLSQLYTP 516

Db 451 PATKLTWITVNMWDLFTAKNNEAFRAPSQMERFVSGAHFGANTGLDHNRFVANNPL-RP 509

Qy 517 QT-----NLDADGKLKPRQGNQF--EVGYKSYMDRLNARVSPFMKDKNAAPL 566

Db 510 ETAKNKEITANLHPDSL F--KQGDFKIEATYFRNDVDQFINLKT--FIDAKTSASAGA 564

Qy 567 NPNNKKTTRAAALGKRMGEVETISGAVVPKMOIHAGSYSLHSQITANSRRDGI FILM 626

Db 565 NPNT-----NGALLPK-----NSQYQNTNARLSGIELQA 594

Qy 627 PKKS--ANLWTTYQVTPBLTIGGVNAMSGITSSAGMHAAGY-----ATPMAAA 674

Db 595 QYQTEBLTFTNGSKTKGDKDQSG-EALSNIAASKIGVGNVYALVMDKFTVGATVTHYAA 653

Qy 675 YRFTPK-----LKQIADNINFRHYARVGANTENIP 707

Db 654 QRVVPRDHSVATYPSYITLDRATVADLKGEMKVLRLDFALENLFDKRYQ-----AFSL 707

Qy 708 -PSERTWTANLRYSF 722

Db 708 MEGTGRNAKISAVYSF 723

RESULT 35

HPUB_NEIMC

ID HPUB_NEIMC STANDARD; PRT; 810 AA.

AC P96949;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)


```

RT Escherichia coli K-12."
RL Biochimie 73:1361-1374 (1991).
CC -1- SIMILARITY TO S. TYPHIMURUM ORF NEAR CYS6 (AC P25928).
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS20/IS90C ELEMENT
CC BETWEEN AMINO ACIDS 839 AND 840.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, AE000237; AAC74483.1; ALT SEQ.
CC EMBL, AE000237; AAC74487.1; ALT SEQ.
CC EMBL, D90778; BAA15009.1; ALT SEQ.
CC EMBL, D90778; BAA18880.1; ALT SEQ.
CC EMBL, D90779; BAA18881.1; ALT SEQ.
CC EMBL, X62680; -; NOT ANNOTATED_CDS.
CC EcoGene; EG11307; ydbA.
CC Hypothetical protein: Complete proteome.
CC CONFLICT 489 I -> V (IN REF. 2).
CC CONFLICT 495 I -> V (IN REF. 2).
CC FT 489 495
CC SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220BE CRC64;

Query Match 3.8%; Score 143; DB 1; Length 2003;
Best Local Similarity 20.9%; Pred. No. 0.12;
Matches 166; Conservative 103; Mismatches 318; Indels 208; Gaps 39;

QY 1 MGQFMSVFRIINMTATVTLAALSSSVFAQAQA-----DLETVHKGQSRYNALIVTEKNG 53
DB 626 VQDFSTGVINGNNNSVTLLAKDLKVVQKATGIVNSGDANTVNTIGNVLDKDKTADNA 685
QY 54 DYSSFAVYTGKIPALREL--PQSVIITNQVKRANVDTEQGLARKTPGLRLVLSNDG 111
DB 686 AEYFPSPGIVNGVGDNNVTLDGKLTIVSDSEVTSKQSLFPGSABKTSGLVVG--DG 743
QY 112 RASVYARGE-YSEVNIIDGLPAQMOSINGTLPFLFAPDRVYRWKRGSGLP-----DSGGE 165
DB 744 NTVNMGGEELIGEKALADGSGVTS---LRTGYSTSYIVVSGSSSVLYNDTTISGE 799
QY 166 ---MGQIVNLVRK-----RPTKAFQGHAAAGFGTHK---QYKAEADVSGSL 205
DB 800 PFLGAGVIRVQDKALLEIGSGATLTMQDIDSPHH---GRTVIEIQLGFAFVTE- 853
QY 206 NSDGSVGRVMACTVQASPRPA-----EKKNRHETFYAAADMDINDPTVLGAGVLYQQRH 260
DB 854 NTTGINSSTISLQNGKDPAPSPVILLATGSGAT-----NAGTITGK---VTEQH 901
QY 261 LAPYNGLPADA-----NNKLPSLPQHVPFQADWPKFKNNSHVDFADIKHYNGGQGVK 314
DB 902 -SVFNKYSITGSTNSPIFNNDVSSITGLV---AQSNSTTINTDGIIDL---YGRGSVMKL 954
QY 315 GMRYSRDR-----ADSNYAFAAGSKGKMTPAKPCNTADPKACAVGLSTETKQKALAF 368
DB 955 ALADSTAEKNGKTLTDSMWWDANDTTAMRDLAS---NSALIDFGTGVGVGTTD----- 1002
QY 369 DASYSRPFPR-----LQNTANEFVIGADYNRFRSTN---EGGRITLYARGGLALNEFRS 418
DB 1003 ---SYSGAGKNAATFAINQLGVTITYNAGAGMAAYGASNTVINQGTINLEKNGY----- 1053
QY 419 IQQVULIANAKGVGRVSHVATENLDERGIY-----GKSTHPADGLSLIGCGRLGHX 472
DB 1054 ---DDSLAATLTVMGMAVYEHGTALN--DQGTVININVGIGQAFYNDGTG--TIVVGTITCF 1108
QY 473 KI-ESGE-----GKTLHKASKTKFTGTGAYGVYL---NDNNSLYLSLSQL 513
DB 1109 GVCQSGNEVYNTDDPTSLITYTGDDTITRSGEYTLNKSAAVTDKLAGNVVNSGTLSGDQI 1168
QY 514 YTPQTNLD-----ADGKLLKPRQNGQFVYGYGYSMD--RLNAAVSYFRMK 558

```

```

DB 1169 TVSSGILENTSGGIIINLYKLDKGAIVIKAGVNTNVVDSGILLNNAAGETAQITM----- 1224
QY 559 DKNAAPLPLN-PNNKTRVYALGRVMEGVETEISGAVTPKQTHAGYSYLSQIKTASNS 617
DB 1225 --NAGDSSLVNNTG-----INKIVNAGVFENNSSGVTRMSAGG-----VFNN 1268
QY 618 RDDGIFLMPKHSANLMTWYQVTPBELTIGGGVNAMSGITSSACMHAAGYATPDMAAYRF 677
DB 1269 QDTGALIM-----RCALLTGTAIVANNECTMNLG-----SSSEGNNTG----- 1304
QY 678 TPKLKQIVADNFIN 692
DB 1305 ---MLEVNNNSAFN 1315

RESULT 37
HXCL_HAEMIN STANDARD; PRT; 744 AA.
ID HXCL_HAEMIN
AC P44523;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin utilization protein C precursor.
GN HKUC OR H10113.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shrivley R., Liu L.-I., Glodex A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512 (1995).
CC -1- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC CONCENTRATIONS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE B OF
CC H. INFLUENZAE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, U32696; AAC21789.1; -.
CC DR TIGR; H10113; -.
CC DR InterPro; IPR000531; TONB boxC.
CC DR Pfam; PR00593; TONB boxC; 1.
CC DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC KW Outer membrane; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC FT SIGNAL 1 24
CC CHAIN 25 744 HEME/HEMOPEXIN UTILIZATION PROTEIN C.
CC SEQUENCE 744 AA; 85043 MW; 74D94F72E1AEC31 CRC64;

Query Match 3.7%; Score 140.5; DB 1; Length 744;
Best Local Similarity 19.5%; Pred. No. 0.046;

```

Matches 162; Conservative 99; Mismatches 269; Indels 301; Gaps 50;

```
QY 41 QRSYNA-----IVTERKNGDYSSFAVTVGTIKIPASLRREIPQSVSIITNQVQKDRNVDTF 93
DB 26 QNSSNSREELLPIIWTND-----SNKLPGR-----SVLKQKNEIQXQADNA 69
QY 94 DQARKTFCGLRVLSN-DDGRSSVYARGVEYSE---YNTDGLPAQWQSI-NGTL---PNLF 145
DB 70 ANLINILPGVNMAGFRPGQGLTNGMGDAEDVRQDGLGATKSPKQYQGSIFTEPELL 129
QY 146 AFDRVEMRGPSGLFDSSGEMGIVNLVRKRPKFAQGHAAAG-----FGTH-----KQVK 196
DB 130 R---RTVDKGNYPQYNGGFGAGTVKFKTKDARDPQENKQIGGLKYGNNSNNQKYS 187
QY 197 ABADYSG-SLNSD-----GSVRGRVMAQTVGASPRP-----AEKNRHETFYAAADWDIN 245
DB 188 TALVLQNEQKNIDLLFGSVR-----NAGDYKRPDNSKILFSKQNO-KTGLIKLNWQIS 240
QY 246 PTVVLGAGVLY-----OORHLAPYNGLPADANNKLPSPLOHVFGAD--WKKFKM 293
DB 241 PELLTLSSVYGIHKGWEPFAAKRDILP-----KPSLSIDIMRYGTDIAW-KRKL 288
QY 294 NSHDVPADLKHYFGNGGKYGKVMYSDRSDADSNYAFAGSKLGMKTPAGRPGCN----- 346
DB 289 ---VTRDQK-----DENYTLKYNL-----PENNPWNLSTQFSY 320
QY 347 ---TADD---KACAVG---LQTE---IKOKALAFDASYSRPRFLNGTANEFVIG----- 388
DB 321 SKTTQNDMRPKASSGLVSLGNSQSWITYSDLTFDINTNSTENIKTTVHELLFGLQWLKN 380
QY 389 -----ADYN-----RPRSTNEGRRTLVARGGLALNEFRSIPQVDLIANA 428
DB 381 TRNTLMYDKSVRKADYNGYFQPYMPSGRQYQAFYLD-----QIKWKNIIFFST 432
QY 429 RGVGVGYSH--TVATENLDEFGIYKSTPHPADGLSLGGRLGHYKTESGSGKTLHKAS 486
DB 433 ---GVR-YDHINNIGQKNL-----ALKYNDISAG-----HDYS 461
QY 487 KTKFTG---YAGAVYDLNDDNSLYLSLSQY-----TPQTNLDADGKL 526
DB 462 QKNGYWSYGLGLNDVNHLSLFTNFSKTRAPVIDEQYEQFKQSSVPATSLNLEKEM 521
QY 527 LKPROGNQEVG-----YKGSYMDRLNARVSFRMKDKNAAPLNPNN 570
DB 522 I-----NOTRVGGIITLNLHLFOENDAFQPRTTYFYNR--GRKEIFKTRGVNCGVNAADTN 574
QY 571 KKT-----RYAALGRKVMGEVETE-----ISGAVTPKWQIHAGSYLHSHOIKTASNS 617
DB 575 NKVCPKIENRNLPQYVIQGALEAYYQSYTLFGEIT-----YSYVKGRQTSFPRN 626
QY 618 RDDGIFLLMPKHSANLWTTYQVTP-----LTIGGGV-----NAMSG- 654
DB 627 -----PWGKTSWIA-EIPPRKATTALGFNVPKYLYTVGWRAEFVRRQDRSPLSGD 676
QY 655 -ITSSAGMHAG-GYATFDMAAYRYTPKLK---LQINADNIENRHYARVG 700
DB 677 PRASSWSLPSARGSYLHNLFLSWS-PAKIKGMNVKITVDNLFNRYNPLYG 726
```

RESULT 38

```
OAR_MYXA
ID OAR MYXA STANDARD; PRT; 1061 AA.
AC P38370;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OAR protein precursor.
GN OAR.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
```

```
RP SEQUENCE FROM N.A., AND SEQUENCE OF 776-796.
RC STRAIN=D2F1;
RX MEDLINE=93328680; PubMed=8335633;
RA Martinez-Canamero M., Munoz-Dorado J., Farez-Vidal E., Inouye M.,
RA Inouye S.;
RT "Oar", a 115-kilodalton membrane protein required for development of
RT Myxococcus xanthus.";
RL J. Bacteriol. 175:4756-4763 (1993).
CC -1- FUNCTION: REQUIRED FOR CELLULAR ADHESION DURING FRUITING BODY
CC FORMATION.
CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED.
CC -1- SIMILARITY: LOCAL, TO TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S64103; AAB27614.1; -.
DR PIR; A40609; A40609.
KW Outer membrane; Signal.
FT SIGNAL 27 1061 POTENTIAL.
FT CHAIN 1 1061 OAR PROTEIN.
SQ SEQUENCE 1061 AA; 114455 MW; EABC077296352EF0 CRC64;
```

Query Match 3.6%; Score 137.5; DB 1; Length 1061;
Best Local Similarity 19.0%; Pred. No. 0.12; Indels 361; Gaps 51;
Matches 187; Conservative 116; Mismatches 321;

```
QY 5 MSVFRINNTAATVLA--LSSSVFAAQTADLETV-----SSFAVTGKTPASLRIPQSV- 77
DB 1 MHLNRVLRQETGVVVAAGLLYGSAAFAQSSTIIGTVIDAQSRQPAADVVTATSPNLQGEQ 60
QY 43 SYNAIVTEKNGDY-----SSFAVTGKTPASLRIPQSV- 77
DB 61 T---VVTDAQNGYRIPOLPPGDYTLRPEKEQPKYARSAIQRLNRTIRNVNELLPEALG 117
QY 78 -----SIITNQVQKDRNV-----TFDQARKTPGLRVLS 107
DB 118 EVVEIVGAPPIDVGSITGMVNVDOEFKRIAVARPGKGATRSFSLAELAFQAG--- 174
QY 108 NDDGRSSVYARGYSEYNIIDL---PAQMOSINGTLNLPFAFDREVEMRGPGLFDSS 163
DB 175 NDNGVSGINGSTSPENGYYVDGLSTNDPA--FGVNASPLSIEFQVDVNIITG--GYMPEF 230
QY 164 G-EMGGIVNLVRKPTKAFQGHAAAGP-----GTHQYKAEADVSGSLSDGSGVRGVM 216
DB 231 GRSTGGVINAVTRSGNEFHGVSFANWTPGTLEGRKQIREGTV-----ITGQNO 281
QY 217 AQTVGASPRPAEKNRHETFYAAADWDINPDTV-LGAGY---LYQORHLAPYNGLP-DA 271
DB 282 LQNLG-----DFGATLGGPILKDKLWFFAGFAPSFTYQHTRLNALRVUDE 328
QY 272 NKLPSLPQHVFGADWNKFKNNSHDVPADLK--HYFGNGGYGKVMYSDRSDADSNYAF 329
DB 329 GNTIKD--ETDFTVAD--AIPGSARKYVADSRTIYQMGKLT-----LIQDNHVSFAL 378
QY 330 AGS-----KLGWKTPEG-----RPGCNTADDKACAVGLGTEIKQKALAFDASYS 373
DB 379 NGTPTSTGLGKLSVNPQSGGLPGVLATRPG-----DFGL-TETKANTTSLALYA 428
QY 374 RPFRLGNTANEFVIGADYNFRST-----NEQGRITL--YAR-----GGLALNEFR- 417
DB 429 GAF-----ADKKVLDANLGNWFHQFQASTLPDGSNLGDRGTGLAGSRMVYTTFRALTLEA 484
QY 418 -----SIPQVDLIANARKGVGYS-----HTVATENLDEFGIYKSTF-----H 456
DB 485 LPEQEGAGCGSTPEQLV---RSPVTGYGVGGPGFMSDQTLDRYOANAKATYLLNALGTH 541
QY 457 -----PADGLSLGGRL-----GHYKIESGEGKTLHKASK-----TKF 490
```

```

Db 542 VKAGVDVLLSPDQVAYGVFFQEGSNYGV-AGQGAHVHARRGYQTGPDSAVTF 600
QY 491 TSYA-----GAYVDL-----NDNNSLYLSQLYTP----- 516
Db 601 TGVATTSTTTGGFLDSDMSIANRVTLNIGRYDVQALYGNDSLLGNGMSPRIGAI 660
QY 517 -----QTNLDADGKLKPRQG-----NOPE 536
Db 661 VDPFANGRAKVEVNFARVEYQVPLNMDRAFCENRISARRSLAEPGQATSCDSSFE 720
QY 537 VQYKSSYMDRLNA-----RVSFYMKAKNAAPLNPNNKTRVYALGKRWEGVETEI 590
Db 721 SQOATCNTSDSNLAIPESSRNVRFTYGTGTVGTPVDPDIK-----AQSSDEIVWGALEYEV 776
QY 591 SGAVTPKMQIHAGYSYLSHQIKTA--SNRDDG--TFLLMPKHS-----ANLWTYQ 638
Db 777 LA-----NTRLGASTHGDMSVIEDMSRDDGNTYFLGNPSGSGFGEPTPRRYNDVNT 830
QY 639 VTPELLTGGVNAMSGITSSA--GMHAGY-----ATPDMAAAYRFTPLKLO 684
Db 831 VYLNRTFADGWLAAQANYTWSRLYGNYPGLFRPEPTQGLDNLISDFDL-----IELL 881
QY 685 INADNI--FNKHYARVAGANTFNI 707
Db 882 ENRTGLPDRTHQIKVFGAKSENI 906

RESULT 39
VYUA_VIRCH STANDARD; PRT; 687 AA.
ID VYUA_VIRCH STANDARD; PRT; 687 AA.
AC Q00964; Q9JQ00;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vibriobactin receptor precursor.
GN VYUA OR VC2211.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
CX NCBI_TaxId=666;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RX MEDLINE=92276356; PubMed=1317381;
RA Butcheron J.R., Stoeber J.A., Payne S.M., Calderwood S.B.;
RT "Cloning, sequencing, and transcriptional regulation of vib, the
RT gene encoding the ferric vibriobactin receptor of Vibrio cholerae.";
RL J. Bacteriol. 174:3729-3738(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RA Liao W.J., Choi M.H., Butcheron J.R.;
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.D., Hatt D.H., Hickey E.K., Peterson J.D., Unanue L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Baes S., Qin H., Drago I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -1- FUNCTION: RECEPTOR FOR FERRIC VIBRIOBACTIN.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; AF030977; AAB86828.1; -
DR EMBL; AB004293; AAF95355.1; -
DR PIR; A41905; A41905.
DR TIGR; VC2211; -.
DR InterPro; IPR000531; TONB_boxC.
DR Pfam; PF00593; TONB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
DR Signal; Receptor; TONB box; Outer membrane; Transport; Iron transport;
DR Complete proteome.
FT SIGNAL 1 37
FT CHAIN 38 687 VIBRIOBACTIN RECEPTOR.
SQ SEQUENCE 687 AA; 76413 MW; 2CB4FA9BBAD1AF87 CRC64;

Query Match 3.6%; Score 137; DB 1; Length 687;
Best Local Similarity 18.4%; Pred. 0.071;
Matches 127; Conservative 71; Mismatches 247; Indels 244; Gaps 27;

QY 18 LAALSSVFAAQTADLETVHIGKRSYNAIVTEKNGDVSSFAVYTGTKIPASLREIPQSV 77
DB 22 LSAAMMGFTGSGFAAYAEQTONTNOEQEMFVL-----VVIKEKTRSIYETSASV 70
QY 78 SIITNQVVDNRVDVFPDQ-LARKTPG-----LAVLSN--DDGRSS--VYARGVEYSY 125
DB 71 EV-----FDQOTERTPGATEIIDDLQILPVLVDSGQSNMPTTRIGDGGP 117
QY 126 NIDGLPAQMSINGTLP-----NLFAFDREVMKRGPSGLFDSSG 164
DB 118 SVGGL--ASTAGNSPRLMNSIDGRSLYSRIARGPRSLMQOYEIYLGQSYIQGN 173
QY 165 EMGGLVNLVRKRPYKPAFGHAAAGFTGHKOYAEADVSGSLNSD----- 208
DB 174 TSAGAIVMKSNPHTHFESAVVAGIGESDYSQTAGMISAPIIDELAFRLSPDQQRDSF 233
QY 209-----GVRGRVNAQTVGA-----SPRAPKNNR----- 232
DB 234 VDLAAFEPAQDPRKLIEMNSVRGKLIYEPALDGFKTTTLTSHMSDRGPOTENINAVAGNEA 293
QY 233-----HETFYAADWDI-----NPDTVLGAGLY-----OQRHLAP 263
DB 294 RRPVYETASFTTAMDIIMHLNLTFFENNLYVADSYDYTPNPSNGDNTDQKEPHIEP 353
QY 264---YNGLPADANNKLPSPLEQHVFGADWNKFKGNSHDVPADLKHYFGNGYGVGNRYS 320
DB 354 LRYIALDGSVN-----TLIGARY--YQSQDDWYIDAASAVPMDGRTKAKSVF-- 400
QY 321 RQADSVYAFAGS-----KLGKTPAGPGCGNTADDKACAVGL----- 357
DB 401--AEVYVALTPSINVLNLAGRFEREQKENVSHPRYKLDVDETSSVFLPLDVAVYPPVQ 458
QY 358-----CTEIKOKALAPDASYSRPFRLGNTANEFVIGAD--YNRPSTNEQORTLYAR 408
DB 459 TYGIRAAKRYNAGSGAGLAPNSMQFTGFRPYEFOGSIINMYEYTRHRFSHSEVLT---- 514
QY 409 GGLALNEFRSIPQVDLIANARKGVRGYSTVATENLDEFGIYK--STFHPADGLSLIG 465
DB 515 -NLFPVDFDSMQMTQOTS-----SGDVFILNDEASTYGAEIGSIVATSSLELF- 563
QY 466 GGRLGHYKIESE--GKTLHAKASKTKFTGYGAAYVDLNDNNSLYLSQLYTPQTNLAD 523
DB 564 -ANLGLKTEFEKTTGNTTELPAPRMSANVGLYDFG----- 600
QY 524 GKLLKPRQGNQF--EVGYKSYMDRLNA 550
DB 601-----QGFESSNAAYTGSIFYSESNS 622

RESULT 40

```



```

Db 1927 KETHEDGVIEYTDG-----FENRRTVTTKD-----GSSKTVAASFIMNQLTKVND 1975
QY 446 EFGYKSTFHPADGL-----SLIGGRLGH-----YKISEGEGKTLKASKTKF 490
Db 1976 ESISYDKNKNRSDGFEYTWDAEDNLTAATKGEDEKPPATYKDE-KGNRIKTYNGKY 2034
QY 491 TSYAGAVALDNDNSLYLSLSQLYTPQNLADADGKLLKPRQNGPEVGYKGYMDRLNA 550
Db 2035 TNY-----FYDGSLSLVLY-----ETDADNNVTK-----SYTYGDSGQLLSTENG 2075
QY 551 RVSFYMKNKNAAPLNPNKKT-----RYAALGRKMEGVETISGAVYPRKQIHAGYISY 606
Db 2076 KKYFHYNAAGDIIAISDSTGKTVAKYQYDAMG-----NPKTEASDEVXDNRYRYAGYQY 2131
QY 607 LHSQIKTAASNRDDGIFLMPK-----SANLMTTYQ 638
Db 2132 -----DETSGLYIMARKYEPNNGVFLSLDPPSGDGLDONGYAYGNNPNVNM 2181
QY 639 VTPE-----LTIGGGVNAMSG-----ITSSAGMHAGYATFD-----AMAAYETPK 680
Db 2182 YDPDGHVWLNVNAGFAVDGYKAYKSGKMGKAWAALANFPGPKIFKASASAYFTTK 2241

RESULT 41
HGBC_HAEIN STANDARD; PRT; 993 AA.
ID HGBC_HAEIN STANDARD; PRT; 993 AA.
AC Q9KIV0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin binding protein C precursor.
GN HGBC.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OK NCBI_taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTM1 N182;
RX MEDLINE=20316037; PubMed=10858226;
RA Cope L.D., Hxkal Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
RT hemoglobin and hemoglobin-haptoglobin binding by nonypeable
RT Haemophilus influenzae.";
RL Infect Immun. 68:4092-4101(2000).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
CC IS REQUIRED FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF221060; AAF80178.1;
CC InterPro; IPR000531; TONB_boxC.
CC Pfam; PF00593; TONB_boxC.1;
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1;
KW Outer membrane; Transport; TonB box; Multigene family; Signal;

```

```

KW Receptor; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 993
FT DOMAIN 26 37
FT REPEAT 26 29
FT REPEAT 30 33
FT REPEAT 34 37
FT SITE 47 54
FT SITE 976 993
SQ SEQUENCE 993 AA; 113616 MW; A551BR3BC641612 CnC64;

Query Match 3.6%; Score 136; DB 1; Length 993;
Best Local Similarity 19.9%; Pred. No. 0.14;
Matches 129; Conservative 76; Mismatches 211; Indels 232; Gaps 33;

QY 5 MSVFRTMTAAATLALSSSV--FAAQTALLETTHIKGRS-----YNAITYENKGY 55
Db 1 MANFRLNLAIVSMGLTGVAAAPTOPTNQPTNQNNSVSEQLQINVSSTNSD- 59
QY 56 SSFAVTGTRIPASLREIPQSVSIITNQO---VKDRNVTDPDLARKTGLRYL-SNDG 111
Db 60 -----IKTPKIAEYTKAKTLERQANNKD-----IVKYEYGVYVEAGRG 103
QY 112 RGSVARGYEYSE--YNIIDLPAMQOSINGTLPLNFAFDRVEVMRGPGLPSSGEMGI 169
Db 104 QSGFAIRGVDENRVAINIDGL-RQAEITLS-----QGFKEFLFEGYGNFNT 148
QY 170 VNLVRKPTKAGQHAAPGTHKQYKADVSGSLNS---DGSYRGVMQATVQASR 225
Db 149 RN-----GAEIETLKEVNTTKGANSIKSGSGSLGSGVYIKTKD---- 186
QY 226 PAEKNNRHETFAAADWDINPPTVLGAGVLYOORHLAPYNGLPADANNKLPSLPQHVFG 285
Db 187 -----ARDYLKND-----YVSYKKGATENNQSPNITL---LA 218
QY 286 ADWNKF-----KNSHDVFPADLKHYFGNGGYKGVKMYSDRDADSNYAFASGKLGMT 338
Db 219 GRYKFDALVWTRRNGHEL-----ENVDYKNADS-----LTGKKR 255
QY 339 PAGRPGCNTADKACAVGLGTEIKQKALAFDASYSRPFRLGNTAN-EFYIGADYRFRST 397
Db 256 EKADP-----YIEQDSTLLKLSFN-----PLENRFTLADLYEHRER 294
QY 398 NEQGRITL-YARG-----LALNEFRSIPQVDL---ANARG 431
Db 295 GQDLSTLLKQSGNETPEVESRHNTDKTKRNNISYSYNSOTPEWDTLKLTYSDRIK 354
QY 432 VR-----GYSHVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGKTL 482
Db 355 TPARDEYCDAGVTHQGTEN-----PTGKLTNG-----KITRNGTDL 394
QY 483 HKASKTFPTGYGA-VYDIN-----DNNSLYLSLSQLYTPQNLDA-----DGKILKR 530
Db 395 Q--FFKKGTAAAGTKYVDPTFDITNQEIVKLTNIGVANDTWYDCLPDCDKGTAKYF 452
QY 531 QGNQPEVGYKGYMDRLNARY-----SFRMDKNAAPLNPNKKT 574
Db 453 EGNA-TYGIIGKRESELETVLNGKRFARIYDK-----DKNNRNR 494

RESULT 42
PAT2_HUMAN STANDARD; PRT; 4349 AA.
ID PAT2_HUMAN STANDARD; PRT; 4349 AA.
AC Q9NT08; Q75091; Q9NSR7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Procadherin Pat 2 precursor (hPat2) (Multiple epidermal growth
DE factor-like domains 1).
GN PAT2 OR MEGF1 OR CDH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20202599; PubMed=10716726;
 RA Wu O., Maniatis T.;
 RT "Large exons encoding multiple ectodomains are a characteristic
 feature of protocadherin genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
 RN [2]
 RP SEQUENCE OF 3777-4349 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98360089; PubMed=9893030;
 RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
 RT "Identification of high-molecular-weight proteins with multiple
 EGF-like motifs by motif-trap screening.";
 RL Genomics 51:27-34(1998).
 RN [3]
 RP SEQUENCE OF 4142-4349 FROM N.A.
 RC TISSUE=restis;
 RA Pouetka A., Wellenreuther R., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 32 CADHERIN DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF231022; AAF61928.1; -;
 DR EMBL; AB011535; BAA32463.1; -;
 DR EMBL; AL157443; CAB75663.1; -;
 DR Genew; HGNC:3596; FAT2.
 DR MIM; 604269; -;
 DR HSP; P00740; 1EDM.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001791; Laminin-G.
 DR Pfam; PF00028; cadherin; 33.
 DR Pfam; PF00008; EGF; 2.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 32.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00282; LamG; 1.
 DR PROSITE; PS00232; CADHERIN_1; 14.
 DR PROSITE; PS00268; CADHERIN_2; 32.
 DR PROSITE; PS00022; EGF 1; 2.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 1.
 KW Transmembrane, Glycoprotein; Repeat; EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 4349
 FT DOMAIN 19 4048
 FT TRANSMEM 4049 4069
 FT DOMAIN 4070 4349
 FT CYTOPLASMIC (POTENTIAL).
 FT CADHERIN 1.
 FT CADHERIN 2.
 FT CADHERIN 3.
 FT CADHERIN 4.
 FT CADHERIN 5.
 FT CADHERIN 6.
 FT CADHERIN 7.
 FT CADHERIN 8.
 FT CADHERIN 9.
 FT CADHERIN 10.
 FT CADHERIN 11.
 FT CADHERIN 12.
 FT CADHERIN 13.
 FT CADHERIN 14.

FT DOMAIN 1561 1758
 FT DOMAIN 1759 1872
 FT DOMAIN 1969 2070
 FT DOMAIN 2071 2171
 FT DOMAIN 2172 2272
 FT DOMAIN 2273 2379
 FT DOMAIN 2380 2481
 FT DOMAIN 2482 2585
 FT DOMAIN 2586 2691
 FT DOMAIN 2692 2797
 FT DOMAIN 2798 2906
 FT DOMAIN 2907 3011
 FT DOMAIN 3012 3113
 FT DOMAIN 3114 3218
 FT DOMAIN 3219 3321
 FT DOMAIN 3322 3426
 FT DOMAIN 3427 3531
 FT DOMAIN 3532 3642
 FT DOMAIN 3643 3773
 FT DOMAIN 3773 3984
 FT DOMAIN 3989 4022
 FT DOMAIN 3297 3302
 FT CARBOHYD 39 39
 FT CARBOHYD 210 210
 FT CARBOHYD 280 280
 FT CARBOHYD 330 330
 FT CARBOHYD 459 459
 FT CARBOHYD 568 568
 FT CARBOHYD 627 627
 FT CARBOHYD 655 655
 FT CARBOHYD 789 789
 FT CARBOHYD 996 996
 FT CARBOHYD 1175 1175
 FT CARBOHYD 1303 1303
 FT CARBOHYD 1383 1383
 FT CARBOHYD 1417 1417
 FT CARBOHYD 1904 1904
 FT CARBOHYD 1998 1998
 FT CARBOHYD 2007 2007
 FT CARBOHYD 2165 2165
 FT CARBOHYD 2183 2183
 FT CARBOHYD 2225 2225
 FT CARBOHYD 2368 2368
 FT CARBOHYD 2387 2387
 FT CARBOHYD 2430 2430
 FT CARBOHYD 2470 2470
 FT CARBOHYD 2547 2547
 FT CARBOHYD 2597 2597
 FT CARBOHYD 2654 2654
 FT CARBOHYD 3125 3125
 FT CARBOHYD 3276 3276
 FT CARBOHYD 3310 3310
 FT CARBOHYD 3430 3430
 FT CARBOHYD 3471 3471
 FT CARBOHYD 3601 3601
 FT CARBOHYD 3772 3772
 FT CARBOHYD 3813 3813
 FT CARBOHYD 3840 3840
 FT CARBOHYD 3873 3873
 FT CARBOHYD 3904 3904
 FT CARBOHYD 3989 3989
 FT CARBOHYD 4117 4117
 FT CONFLICT 4160 4160
 FT CONFLICT 4349 AA; 479383 MW; 6C050CA17AD398ED CRC64;
 SQ SEQUENCE 4349 AA; 479383 MW; 6C050CA17AD398ED CRC64;

Query Match 3.5%; Score 130.5; DB 1; Length 4349;

Best Local Similarity 21.9%; Pred. No. 2.7;

Matches 97; Conservative 52; Mismatches 140; Indels 153; Gaps 21;

OY 358 GTEIKQKALAFDASYSRRPFLGNTANEFVIGADYNRFR--STNEQGRITTLVARGGL---- 411

DB 3234 GTEVLQL-----ATLTP-----GAEKGYRVSGNEQGRFLDARTGILYN 3276

QY 412 ALNERSIPQVDL-IMARKGVNGSH-----TVATENLDEF-GIYKSTFHPA-- 458
 Db 3277 ASLDEFSPKRYFLSIECSKSSSSLDVTTVMVNIDVNEHRPQDPDPYSTLEVALV 3336
 QY 459 -----DG-----LSLGGRLGHYKIESGEG-----KTLHKASKTKF----- 490
 Db 3337 GGVILTVSATDEGRLNSDITLISLIGNOLGHTTHPKKSELQVAKLDRBQASISLKL 3396
 QY 491 -----TGAVAGVYDLDNNNSLYLSLSQVTPQTNLDADQKLK-----P 529
 Db 3397 RATSQGPPLHEDTDIAIQADVNDNPPRFQALNSTTGVGENSPISKVLQLLSDPSP 3456
 QY 530 RQGNQEVGVKGSYMDRLNARVSFRMKDKAAA-PLNPNKTKRYALGKRVMEGVET 568
 Db 3457 ENGPPY-----SFRITKGNNGSAFRVTPDGMVLYAEGLSRBAQEWYQL 3499
 QY 589 EISGA---VTPKQIHAGVSYLHSQIKTASNSDDGFIILMPK-----HSANL- 633
 Db 3500 QIQASDSCGIPPLSSLYSVRVHVTGESHYAPSLPLEFIVTGEDEFQGNVGIHARD 3359
 QY 634 ---WTVYQVTPBELTIG-----GGVNASGSI-----TSAGM----- 661
 Db 3560 PQDTLTVSLAEETLGRHPSVGAPDGIKIAAGLPRGHVSFNTVSDGCTTTAGVHVY 3619
 QY 662 -HAGVATFDA--MAAYRFTPK 680
 Db 3620 WHVGEALQOAMMMGFYOLTP 3641

RESULT 43

CH11_BACCI STANDARD; PRT; 699 AA.

ID CH11_BACCI
 AC P20533;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Chitinase A1 precursor (EC 3.2.1.14).
 GN CH1A1.
 OS Bacillus circulans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxId=1397;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WL-12;
 RX MEDLINE=90368776; PubMed=2203782;
 RA Watanabe T., Suzuki K., Oyanagi W., Ohnishi K., Tanaka H.;
 RT "Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed
 its evolutionary relationship to Serratia chitinase and to the type
 RT III homology units of fibronectin.";
 RL J. Biol. Chem. 265:15659-15665(1990).
 RN [2]
 RP MUTAGENESIS.
 RC STRAIN=WL-12;
 RX MEDLINE=93366760; PubMed=8103047;
 RA Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H.,
 RA Uchida M., Tanaka H.;
 RT "Identification of glutamic acid 204 and aspartic acid 200 in
 chitinase A1 of Bacillus circulans WL-12 as essential residues for
 RT chitinase activity.";
 RL J. Biol. Chem. 268:18567-18572(1993).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
 acetyl-D-glucosamine polymers of chitin.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
 HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M57601; AAA81528.1; -
 DR PIR; A38368; A38368.
 DR HSSP; P02751; IFNA.
 DR InterPro; IPR003610; CBM_5_12.
 DR InterPro; IPR001579; Chitinase_18/2.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003962; FNIII_repeat.
 DR InterPro; IPR01223; Glyco_hydro_18.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00704; Glyco_hydro_18; 1.
 DR Pfam; PF02839; CBM_5_12; 1.
 DR PRINTS; PR00014; FNTYPEIII.
 DR ProDom; PD000471; Glyco_hydro_18; 2.
 DR SMART; SM00495; CatBD3; 1.
 DR SMART; SM00060; FN3; 2.
 DR PROSITE; PS01095; CHITINASE_18; 1.
 KW Hydrolyase, Glycosidase, Chitin degradation, Signal, Repeat.
 FT SIGNAL 1 41
 FT CHAIN 42 699
 FT DOMAIN 42 460 CATALYTIC.
 FT DOMAIN 465 549 FIBRONECTIN TYPE-III (R-1).
 FT DOMAIN 560 644 FIBRONECTIN TYPE-III (R-2).
 FT ACT_SITE 204 204 PROTON DONOR (PROBABLE).
 FT ACT_SITE 200 200 D->N: DECREASE IN ACTIVITY.
 FT MUTAGEN 200 200 D->E: NO CHANGE IN ACTIVITY.
 FT MUTAGEN 204 204 E->D, Q: LOSS OF ACTIVITY.
 SQ SEQUENCE 699 AA; 73677 MW; AC7C9B2E2987643 CRC64;

Query Match 3.4%; Score 129; DB 1; Length 699;
 Best Local Similarity 21.1%; Pred. No. 0.26;
 Matches 127; Conservative 47; Mismatches 208; Indels 220; Gaps 27;

QY 198 EADVSGLSNDSQSVRRWAAQTVGASPRPAEKNNR-----HETVYAADMDINDTVL 250
 Db 204 EYVSGGLD-----GNSKRPEDDKQNTYLLSKREKLDAGAVD----- 242
 QY 251 GAGYLYQOQHLPYNGLPAD--ANNKLPSLPQHVFVADAMNKKFNSHVFADLKHYFGN 308
 Db 243 GKTYL-----LTIASGASNTYAMTELAKIAIV-----DW-----INIMTYDNR 282
 QY 309 GGYKVMRYSDRDADSNYAFAGSKLGKMTKPAKPGPCNTADKACVGLTEIKQALAF 368
 Db 283 GAMQKI-----SAHNAPLNDPAAS-----AAGVPDANFNVAAGAQ-----HL 322
 QY 369 DASYSRFLGNTANSPVIGADYNNRPSRINTEGRTLLVARGGLALNERSIPOVDLIANA 428
 Db 323 DA-----GVPAALVLGVPP-----YGRGW-----DGCQAQ 348
 QY 429 RKGVRSHTVATENLDEFGIYKSTFHPADGLSLGGRLGHYKIESGKTLHKASKT 488
 Db 349 -----GNGYQTCGTGSSSVGTWEASFPDYDEANYNKNGYT 386
 QY 489 KFTGVAGAV-YDLDNNNSLYLSLSQVTPQTNLDADGKLLKPRQGNQFEVGYKGYMDR 547
 Db 387 RYWNDDAKVPLYLNASNKRKFSYD-----DAE-----SVGYTAVIKSK 425
 QY 548 LNARVSFYMK-DKN-----AAAPLPNNKKT----- 573
 Db 426 GLGAMFWELSGDRNKTLLONKLAADLPTGTVPPVDTTAPSVFGNARSQVTANSVTLAM 485
 QY 574 -----RYALGKRVMEG--VETELSGAVPKQIHHGVSYLH-----SQIK 612
 Db 486 NASTDVGVTGVNVNNGALASVITGTTATISGLTGTGTYTPTIKAKDAAGNLASAANAV 545
 QY 613 TASNHRDDGIFILMPKHSANLWTVYQVTPBELTIG-----GGVNASGITSAGNH 662
 Db 546 TVSTTQPGDQAPAPAPNLASTAGTSSITLSWTASTDNGVGVGYNG-TALATTV 604
 QY 663 AGGYATFDMAA---YRFTPKLQINADNINRHHYAAVGAGANTNIGSERTWTANLR 719
 Db 605 TGTATISGLADDTSTYTFVAK--DAAGNVASAASNAVSVKTALETNTNGVS-AWQVNTA 661

DE (Sucrose 6-glucosyltransferase).
 GN GTFD.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxId=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GS-5;
 RX MEDLINE=91100958; PubMed=2148600;
 RA Honda O., Kato C., Kuramitsu H. K.;
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
 the glucosyltransferase-S enzyme."
 RL J. Gen. Microbiol. 136:2099-2105(1990).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MT4239, MT4251, MT4467, and MT8148;
 RX MEDLINE=98231643; PubMed=9570124;
 RA Fujimura T., Teruo Y., Hoshino T., Kawabata S., Osobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 Streptococcus mutans."
 RL FEMS Microbiol. Lett. 161:331-336(1998).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
 fructose + {(1,6)-alpha-D-glucosyl}(N+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES
 WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
 FORMS OF GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 BINDING PROTEIN FROM S. MUTANS.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use. By non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, M29286; AAA26895.1; -
 DR EMBL, D88653; BAA26103.1; -
 DR EMBL, D88656; BAA26107.1; -
 DR EMBL, D88659; BAA26111.1; -
 DR EMBL, D88662; BAA26115.1; -
 DR EMBL, D89979; BAA26121.1; -
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; GH_70.
 DR Pfam; PF01473; CW_binding_1; 11.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Transferrase; Glycosyltransferase; Signal; Repeat; Dental caries.
 FT SIGNAL 1
 FT CHAIN ? 1462
 FT DOMAIN 1232 1423 GLUCOSYLTRANSFERASE-S
 FT REPEAT 1232 1295 3 X 63 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 1296 1359 1.
 FT REPEAT 1360 1423 2.
 FT VARIAT 58 58 3.
 FT VARIAT 68 68 K -> E (IN STRAIN MT4467).
 FT VARIAT 81 81 A -> S (IN STRAINS MT4239 AND MT4245).
 FT VARIAT 113 113 A -> T (IN STRAIN MT4251 AND MT8148).
 FT VARIAT 122 122 T -> I (IN STRAINS MT4239 AND MT4245).
 FT VARIAT 122 122 A -> V (IN STRAINS MT4239, MT4245 AND
 MT8148).
 FT VARIAT 132 132 S -> A (IN STRAINS MT4239, MT4245, MT4251
 AND MT8148).
 FT VARIAT 135 135 A -> V (IN STRAIN MT4245).
 FT VARIAT 202 202 V -> L (IN STRAIN MT4239).

FT	VARIANT	255	255	D -> N (IN STRAIN MT8148).
FT	VARIANT	275	275	E -> D (IN STRAINS MT4239, MT4245 AND MT4251).
FT	VARIANT	288	288	D -> N (IN STRAINS MT4239, MT4245 AND MT4251).
FT	VARIANT	301	301	Q -> H (IN STRAIN MT4245).
FT	VARIANT	313	313	D -> N (IN STRAINS MT4239 AND MT4251).
FT	VARIANT	317	317	E -> K (IN STRAIN MT4239).
FT	VARIANT	328	328	V -> F (IN STRAIN MT4239).
FT	VARIANT	350	350	F -> L (IN STRAINS MT4239, MT4251 AND MT4467).
FT	VARIANT	628	633	KKKYQ -> EKEYTL (IN STRAIN MT4251).
FT	VARIANT	688	688	A -> S (IN STRAIN MT4239).
FT	VARIANT	726	732	TDOGSEA -> ADKNDK (IN STRAIN MT4251).
FT	VARIANT	726	730	TDOGS -> ADKGN (IN STRAINS MT4239 AND MT4245).
FT	VARIANT	964	964	D -> Y (IN STRAIN MT4251).
FT	VARIANT	1019	1019	E -> K (IN STRAINS MT4245 AND MT4251).
FT	VARIANT	1059	1060	LG -> IR (IN STRAIN MT4251).
FT	VARIANT	1060	1060	G -> R (IN STRAIN MT4245).
FT	VARIANT	1080	1080	G -> R (IN STRAIN MT4239).
FT	VARIANT	1142	1142	O -> H (IN STRAINS MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT	VARIANT	1198	1198	S -> N (IN STRAIN MT4239).
FT	VARIANT	1220	1220	Y -> C (IN STRAINS MT4251 AND MT4467).
FT	VARIANT	1280	1280	F -> L (IN STRAIN MT4467).
FT	VARIANT	1282	1282	Q -> P (IN STRAIN MT4245).
FT	VARIANT	1290	1290	K -> T (IN STRAIN MT4245).
FT	VARIANT	1311	1311	N -> D (IN STRAIN MT4245).
FT	VARIANT	1403	1403	D -> G (IN STRAINS MT4239, MT4245, MT4251 AND MT8148).
FT	VARIANT	1425	1425	R -> G (IN STRAINS MT4239, MT4245, MT4251, MT4467 AND MT8148).
FT	VARIANT	1449	1449	R -> K (IN STRAIN MT4467).
FT	CONFLICT	1428	1462	RYDKNSGMYNKVITLNGRRIGIDRWGLARY -> VYR (IN REF. 1).
SEQ	SEQUENCE	1462 AA;	163512 MW;	5C6541FDDCB0DP00 CRC64;

Query Match 3.4%; Score 129; DB 1; Length 1462;
 Best Local Similarity 18.3%; Pred. No. 0.73; Indels 290; Gaps 40;
 Matches 138; Conservative 104; Mismatches 221;
 14 AATVLAALSSVFAAQT-DLETVHIKQGRVNAITKNGYSSPAVVGTKIPASLRE 72
 838 AASNKANMAGQVYESSALDSQILY-EGFSNQDPFT-KDSYTN-----KK 882
 73 IPQSVIITNQVKDRNVDTFQDLARKTPGLRVLSNDDR--SSVYARGEYSEYNIDGL 130
 883 IAGNV-----QLFKSWGVTSEMAP-----QVSESDGSFLDSITONGVARE-- 924
 131 PAOMOSINGTLNLPAPFDEVEVMKRGSSGLPDSSEGGGLVNLVRRK-----PTKAF 181
 925 -----DYYDLAMSKNNRYGSQODMINAKLHSGIOVIADWVDDQIY 967
 182 -----QGHAAAGFTHQKKAADVSGSL-----NSDG-----SVRGVMAQVSGSPR- 225
 968 NUGKEVATRTAVNDVYGEFRKSKSEIKTYLAANTSGNDYOAKYCGAATLSLAATYPSI 1027
 226 -----PAEKNNRHETFYAAADWDINPDTVL--GAGYLYOQHRLAPY----- 264
 1028 FNRQTQSNKKIDPSBKRTAMAKAY-----FNGVTIILGRGVYVLKDNASDKYFELKGN 1081
 265 -NGLPADANNKLPSSL-----PQHVFP--GADNNKRNKNSHVF-- 299
 1082 QTYLPRQMTNKAASGTGVNDGNGMTFYSTSGYQAKNSFQADAKKMWYFDNNGHMYGLQ 1141
 300 -----ADLKHYFGNG-----GYGKVGKRYSDR-----DADSNYAFAGSK-- 333
 1142 QUNGEVQVYLSNGVQLRESFLENADGSKRYFGLHGRYNGYVSDNDNSKMWYFPDASGM 1201
 334 -LGMKTPAGRPCNMTADDKACAVGLGTEIKOKALAFDASYSRPFRLGNTANEFVIGADYN 392
 1202 AVGLKTINGNTQVFPQD-----GYQVK-----GAWITGSD-- 1231

Search completed: December 18, 2002, 06:42:58
Job time : 29.3071 secs